Biol. Cybernetics 33, 9–28 (1979)

Biological Cybernetics

A Theory for the Acquisition and Loss of Neuron Specificity in Visual Cortex*

Leon N Cooper, Fishel Liberman, and Erkki Oja[1]

Center for Neural Science and Department of Physics, Brown University, Providence, R.I., USA

Abstract. We assume that between lateral geniculate and visual cortical cells there exist labile synapses that modify themselves in a new fashion called threshold passive modification and in addition, non-labile synapses that contain permanent information. In the theory which results there is an increase in the specificity of response of a cortical cell when it is exposed to stimuli due to normal patterned visual experience. Non-patterned input, such as might be expected when an animal is dark-reared or raised with eyelids sutured, results in a loss of specificity, with details depending on whether noise to labile and non-labile junctions is correlated. Specificity can sometimes be regained, however, with a return of input due to patterned vision. We propose that this provides a possible explanation of experimental results obtained by Imbert and Buisseret (1975); Blakemore and Van Sluyters (1975); Buisseret and Imbert (1976); and Frégnac and Imbert (1977, 1978).

Introduction

Experimental work of the last generation, beginning with the pathbreaking work of Hubel and Wiesel (1959, 1962), has shown that there exist cells in visual cortex (areas 17, 18 and 19) of the adult cat that respond in a precise and highly tuned fashion to external patterns – in particular bars or edges of given orientation and moving in a given direction. Much further work (Blakemore and Cooper, 1970; Blakemore and Mitchell, 1973; Hirsch and Spinelli, 1971; Pettigrew and Freeman, 1973) has been taken to indicate that the number and response characteristics of such cortical cells can be modified. It has been observed in particular by Imbert and Buisseret (1975); Blakemore and Van Sluyters (1975); Buisseret and Imbert (1976); and Frégnac and Imbert (1977, 1978), that the relative number of cortical cells that are highly specific in their response to visual patterns varies in a very striking way with the visual experience of the animal during the critical period.

These results provide strong evidence for the modification by experience of the response characteristics of individual cortical neurons. A question of great interest is just what the form of such modification is. In this paper, we show that a proposed new form of neuron (synaptic) modification, closely related to prior forms that have been shown to lead to distributed memories in neural networks, can also account for many of the experimental results.

Imbert and Buisseret have classified cortical cells that respond to visual stimuli into three groups – aspecific, immature and specific.They and Frégnac and Imbert have measured the relative proportions of these groups depending on the visual experience of the animal.

The classification of cells used was as follows:

Non-specific Units

The receptive fields of these cells are usually circular and often very large (5–15°). They are activated equally well by bars or spots of light moving in any direction over their receptive field.

Immature Units

The receptive fields of these cells tend to be rectangular (10° × 8°) and the response to rectilinear stimuli is greater than to spots. When moving stimuli are used, their selectivity as a function of the direction of movement is greater when bars or slits are used than with spots. They exhibit a degree of orientation selec-

* This work was supported in part by a grant from the Ittleson Foundation, Inc.

[1] *Permanent address*: Department of Technical Physics, Helsinki University of Technology, SF-02150 Espoo 15, Finland

0340-1200/79/0033/0009/$04.00

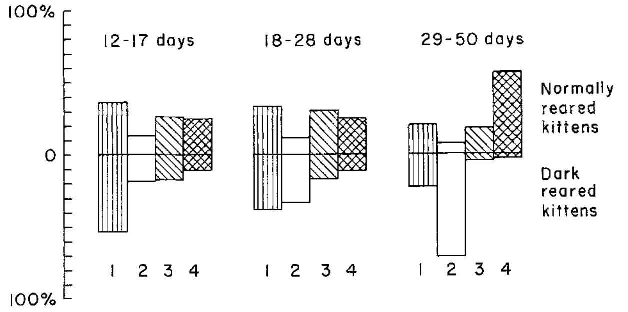

Fig. 1. Distribution of the different types of cells in three age groups in the normally reared kittens (upper part) and in the dark-reared kittens (lower part), from Frégnac and Imbert (1977, 1978). We have normalized the ordinate so that the heights are the percentages of cells in the various functional groups. Type 1, non-activatable (▥); 2, non-specific (□); 3, immature (▨); 4, specific (▩)

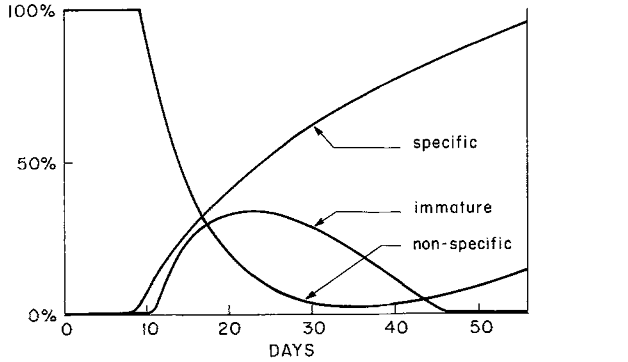

Fig. 2. Evolution of the development of the various specificity groups in cats raised normally. The curves taken from Y. Frégnac (1978) are the results of his unweighted regression analysis of experimental data on 1050 cells

tivity although responding to a range of orientation up to 45° either side of the optimal orientation. In all cases there is a clear null orientation, orthogonal to the optimal, which has no effect.

Specific Units

These cells are orientation-specific and exhibit all the characteristics of simple or complex cells in the adult cat (Hubel and Wiesel, 1962; Henry et al., 1974; Pettigrew, 1974). Such cells exhibit sharp tuning curves and do not respond to orientations more than 30° on either side of the optimal orientation. In addition, their receptive fields are smaller in size (2° × 4°) than those of immature cells.

The distribution of the different cell types in three age groups is shown in Figure 1.

Examination of these results, which were obtained from the study of 1050 cells, confirms that cells having some of the highly specific response properties of adult visual cortical neurons, especially concerning orientation selectivity are present in the earliest stages of post-natal development independent of visual experience (Frégnac and Imbert, 1977, 1978). However, visual experience between 17 and 70 days is critical in determining the evolution of these cells. Animals reared normally showed a marked increase in the number of specific cells as compared with aspecific. (The period between 17 and 28 days is usually sufficient to reach the normal adult level of specificity.) The reverse is true for animals reared in the dark. A statistical analysis of this evolution, performed by Frégnac, (1978) shows clearly the striking dependence of the ratio of sharply tuned to broadly tuned cells depending on the experience of the animal (Figs. 2 and 3).

In addition as has been shown by Imbert and Buisseret (1975); Buisseret and Imbert (1976); and Buisseret et al. (1978) as little as six hours of normal visual experience at about 42 days of age can alter in a striking fashion the ratio of specific or immature to aspecific cells (Fig. 4). That such a short visual experience can change the tuning ratios so markedly is clear evidence of the great plasticity of these cortical cells at the height of the critical period.

These results seem to us to provide direct evidence for the modifiability of the response of single cells in the cortex of a higher mammal according to its visual experience. Depending on whether or not patterned visual information is part of the animal's experience, the specificity of the response of cortical neurons varies widely. With normal patterned experience specificity, developing normally, increases. Deprived of normal patterned information (dark-reared or lid-sutured at birth, for example) specificity decreases. Further, even a short exposure to patterned information after six weeks of dark-rearing can reverse the loss of specificity and produce an almost normal distribution of cells.

The data also indicate that for some cells, at least, some orientation preference is built-in and develops independent of very early visual experience. This initial preference seems to be quite stable in acute experiments (Bienenstock and Frégnac, to be published) and further may be stable under limited dark-rearing conditions since orientation preference can be retrieved with visual experience. There is some evidence however that the orientation columns that result after initial deprivation of patterned experience are not necessarily the same as those that result with normal experience (Blakemore and Van Sluyters, 1974; Movshon, 1976).

In what follows we present a theory (which generalizes the theory of Nass and Cooper (NC), 1975) in which we attempt to account for many of these results.

I. Mapping from the Visual Field to Visual Cortex

The organization and receptive fields of retinal ganglion cells as well as the relay through the lateral

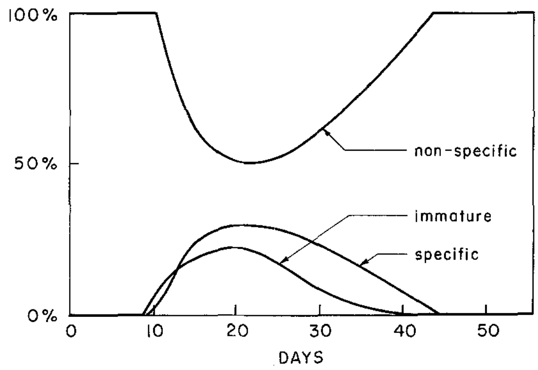

Fig. 3. Evolution of the development of the various specificity groups in cats raised in total darkness. The curves taken from Y. Frégnac (1978) are the results of his unweighted regression analysis of experimental data on 1050 cells

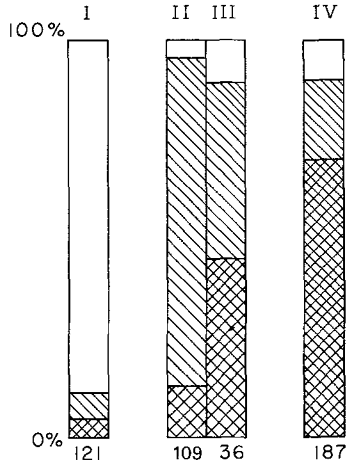

Fig. 4. Distribution in % of the three types of visual cortical units (area 17) recorded after visual exposure in six week old dark-reared kittens, from Buisseret et al. (1978). Visual stimuli, slits or spots, were displayed on a tangent screen. Columns: I, dark-reared, IV, normally reared kittens. During 6 h of exposure, conditions were: II and III, freely moving; in III, 12 h in the dark followed the 6 h of exposure. Numbers of visual cells recorded are given under each column. Specific cells (▩) are activated by orientated stimuli within a sharp angle ($<60°$). Immature cells (▧) are activated by orientated stimuli within a larger angle ($<150°$). Non-specific cells (□) are activated by non-orientated stimuli moving in any direction. A statistical analysis reveals no significant difference in the percentages of immature and specific units between columns III and IV. Therefore it may be that a six hour exposure to visual input followed by twelve hours in the dark is sufficient to produce a distribution of cortical cells similar to that of normally reared animals

geniculate nucleus (LGN) have been intensively studied and the neural pathway from retina to visual cortex is one of the best known signal transmitting systems in a brain. For our present purpose, however, the precise details of this mapping are not as important as the requirement that there exist a correspondence that is sufficiently one-to-one, between external stimuli in the visual environment and the actual sensory input to the cortical neuron. This is a basic assumption in what follows.

In a visually active animal, the stimuli converging on a cortical neuron at a given time, are predominantly determined by (or have as a dominating component) the visual pattern falling on the retina. We generalize the mapping of Nass and Cooper (1975) in which a pattern e^k in the external world is mapped into activity in cortical cells to include two classes of synaptic junctions between cortical and lateral geniculate cells: those which are modifiable and those which are not at all (or substantially less) modifiable. (A similar hypothesis has been made by Blakemore and Van Sluyters (1975).) We do this to include the possibility that innate instructions are contained in non-modifiable junctions. This results, under circumstances that will become clear later, in the ability of a cortical cell to retrieve specificity to its original orientation preference even if it has lost this specificity due to lack of patterned visual input during some portion of the critical period. (The possibility of having innate information in non-modifiable junctions that already existed in NC is also retained.)

We thus divide stimuli carried by input fibers to a cortical cell into two groups, depending on the modifiability of the junctions through which they are transmitted to the post-synaptic membrane. These stimuli are represented as two simultaneous spatial pattern vectors, one comprised of those signals that are input to modifiable junctions, the other comprised of those signals that are input to weakly modifiable, or non-modifiable junctions as shown in Fig. 5. A pattern e^k in the external world impinges on the retina and is transmitted through the retinal-LGN pathway, evoking a set of responses at the outputs of LGN cells to visual cortex.

These responses are the elements of two signal vectors b^k and d^k which then are input into layer IV C of visual cortex to cortical neurons sharing approximately the same receptive field, the one containing the pattern e^k.

It should be stressed that the vectors b^k and d^k are labelled by definition, according to the modifiability of their respective synaptic junctions with the cortical cell. The elements of both vectors are just the responses in different LGN axons to one and the same external stimulus e^k. It is therefore possible that these two vectors are equal (which would be the case, for example, if the two types of synapses share the same input fibers) or that one or the other is zero, corresponding to a situation in which one or the other type of synapses is missing. It is also possible that modifiable or non-modifiable synaptic junctions lie close together and have similar basic properties (excitatory, inhibitory) or it might be that they are separated and have

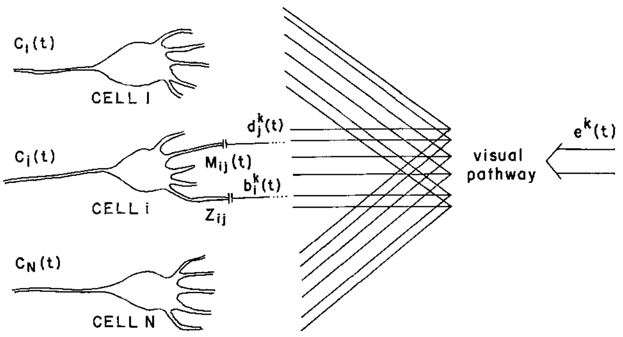

Fig. 5. A visual pattern $e^k(t)$ (e.g., a bar or an edge of a given orientation and moving in a given direction) falls on the retina at time t. This pattern evokes a set of parallel responses in the axons of the LGN cells, and these are presynaptic inputs to a layer of modifiable neurons in visual cortex. The inputs are divided in two parts in the model: vector $d^k(t)$ contains $d^k_j(t)$ that fall on modifiable junctions, vector $b^k(t)$ contains $b^k_j(t)$ that fall on non-modifiable junctions. The strengths at time t of the modifiable junctions, $M_{ij}(t)$, comprise a matrix $M(t)$, while the strengths of the non-modifiable junctions, Z_{ij}, comprise a matrix Z. The responses of cortical neurons, making up a vector $c(t)$, are obtained from weighted sums of the different inputs with weights given by the corresponding junction strengths $M_{ij}(t)$ and Z_{ij}

different properties. (For example, the modifiable junctions might be purely excitatory.)

The response of a given cortical neuron depends on the combined potentials produced by both b^k and d^k. Assuming linearity, the effects of these individual stimuli can be expressed using two junctions matrices, here called M and Z, instead of just one as in NC. The elements M_{ij} of the i^{th} row of M denote the strengths of the modifiable junctions of the i^{th} cortical neuron, and the elements Z_{ij} of the i^{th} row of Z denote in a similar fashion, the non-modifiable junction strengths of that same neuron. The basic component of the post-synaptic potential, produced by the external stimulus e^k, on the soma of the i^{th} cortical cell then is:

$$\sum_{j=1}^{L_d} M_{ij}d^k_j + \sum_{j=1}^{L_b} Z_{ij}b^k_j .$$

Here, using the same notational rules as NC, d^k_j and b^k_j are the j^{th} element of d^k and b^k, respectively. L_d and L_b are the dimensions of d^k and b^k or the numbers of columns in the matrices M and Z, and need not be equal. It is also likely that the number of the cortical neurons, N, (the number of rows in both matrices M and Z) is different from either L_d or L_b. The mutual ratios of N, L_d, and L_b depend on the amount of convergence in the mapping from LGN to visual cortex, and on the relative numbers of modifiable vs. non-modifiable junctions on the cortical neurons.

The external stimulus e^k is not the only factor producing activity in cortical cells. Even in the absence of patterned external stimuli, as is the case, for example, when the eyes are covered, there is neural activity. This is due in part to variations in spontaneous activity in the presynaptic pathways, dark discharges from the retina, input from the reticular formation or other such inputs that are non-specific with respect to visual stimuli (Batini and Buisseret, 1974; Buisseret et al., 1978). For simplicity, these will all be put together to form an additional component x, that can be characterized as channel noise which, in this paper, will be taken to be a random fluctuation about the post-synaptic passive potential. We then have for the post-synaptic potential on the soma of the i^{th} cortical cell, denoted by c^0_i:

$$c^0_i = \sum_{j=1}^{L_d} M_{ij}d^k_j + \sum_{j=1}^{L_b} Z_{ij}b^k_j + x \tag{1.1}$$

Channel noise should be distinguished from another possible random component in the input, whose sources are variations in the actual input pattern e^k. In a normal visual environment, optimal patterns will not usually appear alone on the retina: they will generally be immersed in a background that is likely to vary from one realization to another.

If e^k is given a time index $e^k(t)$, where it is understood that $e^k(t)$ stands for the external pattern actually occurring at time t, in which the optimal pattern e^k is a prominent component, then d^k and b^k will also become dependent on time and will consist of an optimal pattern and a component (that will be shown to be noise-like in the next section) as follows:

$$\begin{aligned} d^k(t) &= d^k + r(t) \\ b^k(t) &= b^k + s(t) \end{aligned} \tag{1.2}$$

The constant patterned parts of $d^k(t)$ and $b^k(t)$ continue to be written d^k and b^k and, as will be shown, the noise vectors $r(t)$ and $s(t)$ can be taken to have zero-mean.

It is generally agreed that the input-output function by which a cortical neuron maps dendritic post-synaptic potentials into axon firing rates is approximately linear in the central part of the frequency scale, and cutoff at upper and lower ends. At the lower end, where the combined potential from all input channels is small, there is a threshold due to the inability of the input to raise the membrane potential to a firing level. At the upper end, a natural saturation limit in the output frequency is imposed by the refractoriness or "dead time" of the neuron membrane.

Measuring the post-synaptic potential from that required to produce the mean spontaneous firing rate, let θ_F be the post-synaptic potential at the firing threshold and let μ be that corresponding to the saturation limit. We let all proportionality constants be absorbed in θ_F and μ in such a way that it is meaningful to compare the input frequencies d^k_j and b^k_j

weighted by the junction strengths M_{ij} and Z_{ij}, to θ_F and μ. Now changing the conventions of NC somewhat define P (essentially the input-output function of a cortical cell) as follows:

$$\begin{array}{lll} & =\mu & \text{if} \quad u>\mu \\ P(u) & =u & \text{if} \quad \theta_F \leqq u \leqq \mu \\ & =\theta_F & \text{if} \quad u<\theta_F \end{array} \tag{1.3}$$

Collecting the above yields for the actual instantaneous firing rate c_i of the i^{th} cortical neuron (in the absence of lateral inhibition or excitation from other neurons)

$$c_i(t)=P[c_i^0(t)]=$$

$$P\left[\sum_{j=1}^{L_d} M_{ij}(t-1)d_j^k(t)+\sum_{j=1}^{L_b} Z_{ij}b_j^k(t)+x(t)\right] \tag{1.4}$$

at a time t when an image of the external pattern $e^k(t)$ appears on the retina.[1]

In (1.3) and (1.4) both the input and the output are given on a scale in which *zero level means the mean spontaneous firing frequency*. So inputs $d_j^k(t)$ and $b_j^k(t)$, as well as output $c_i(t)$, may be negative or positive. The random component $x(t)$ represents fluctuations in the spontaneous frequency and gives a measure to the amount of variance in it. If spontaneous activity is larger than zero, the firing threshold, θ_F, is negative. In the above, one difference as compared to previous models, is the appearance of the term containing non-modifiable junction strengths Z_{ij}. This makes possible a reversible behavior that shows a close resemblance to experimental results.

II. The External World as Seen by a Cortical Cell

Some visual cortical cells are known to respond preferentially to edges or bars of given orientation in the earliest stages of post-natal development, independent of visual experience. In addition, cortical cells are organized in orientation columns so that cells in a single column will respond preferentially to a bar or an edge of given orientation in the visual field (possibly exerting a mutually excitatory effect on one another).

Although it is likely that all orientation preferences occur among cortical cells, for a normally reared animal, orientation preferences shift discontinuously by about 15° from one orientation column to the next. These shifts are seen in electrode penetrations in the plane perpendicular to the column direction (Hubel and Wiesel, 1959). It thus seems reasonable to suppose that interacting orientation columns, those corresponding to a single mapping (M, Z), correspond to discrete orientations separated by about 15°.

We therefore assume that in the orientation preference space there are K basic patterns $e^1, e^2, \ldots, e^K$, corresponding, for example, to bars of differing fixed orientations moving across the receptive field of the cortical cell. A typical value used in our simulations is $K=7$. (This could correspond to a situation in which the orientation directions differ from one another by 15° angles and cover the full circle if we assume that the tuning curves of cortical neurons are symmetrical around their peak orientation and have no directional preference.)

One way in which innate orientation preferences can be built into the mapping from lateral geniculate to cortical cells is as follows: Suppose that initially the contribution of the modifiable junctions can be neglected and further that Z has the form

$$Z=\theta_M \sum_{k=1}^{K} c^k x b^k. \tag{2.1}$$

here $b^1 \ldots b^K$ are the signal vectors in the LGN pathway, input to the stable junctions, corresponding to $e^1 \ldots e^K$, and c^k denotes a unit vector referring to the k^{th} cortical neuron

$$c_j^k=\delta_{jk}.$$

θ_M is the modification threshold, whose meaning will be explained in Section III. Z has the form of a widely tuned correlation matrix memory as has been employed, for example, by Anderson (1970, 1972) and Kohonen (1972). In the absence of noise the passive potential on the body of the i^{th} cell due to the input pattern pair (d^i, b^i) corresponding to e^i is:

$$(c^i, Md^i+Zb^i) \simeq \theta_M \tag{2.2}$$

We assume that the vectors b^i and d^i are normalized. Any other pattern pair (d^l, b^l) would give

$$(c^i, Md^l+Zb^l) \simeq \theta_M(b^i, b^l) \tag{2.3}$$

which in general would be smaller than θ_M if the patterns do not overlap too much so that $(b^i, b^l)<1$.

[1] In analogy with common practice in systems theory, we choose to define our output at a given instant as resulting from the input at the same instant and the "state" (M, Z) at the preceding instant. It may be argued that this is not quite correct; in NC, $\gamma M(t-1)$ was used instead of $M(t-1)$ in equations corresponding to (1.4), to reflect the fact that there is in fact a continuous decay going on during the integration period that is represented by one discrete step in the algorithms. However, since here and in what follows $M(t-1)$ is multiplied by $d^k(t)$, we can simply change the definition of $d^k(t)$ vectors so that γ will be absorbed in them. This will now cause a redefinition of the modification parameters of Section III. Since the modification algorithm presented there is quadratic in $d^k(t)$, the extra γ must be thought to be absorbed in the modification strengths η^+ and η^-. If the "old" values, corresponding to the notation in NC, are $\hat{d}^k(t)$, $\hat{\eta}^+$ and $\hat{\eta}^-$, then the "new" ones employed in the present paper are $d^k(t)=\gamma\hat{d}^k(t)$; $\eta^+=1/\gamma\hat{\eta}^+$; $\eta^-=1/\gamma\hat{\eta}^-$. The substitution above is implicit in everything that follows. As $0<\gamma \leqq 1$, none of these redefinitions cause any mathematical problems

Thus in the absence of noise, cell i will respond preferentially to pattern e^i. This preferential response could be enhanced by excitation from other cells in the same orientation column and by inhibition from cells in neighboring columns of different orientation preference[2].

In normal visual experience optimal patterns such as straight lines or edges seldom appear alone; they are generally immersed in a highly variable environment that is extremely difficult to characterize in any satisfactory fashion. A normally reared animal encounters an immense variety of patterns and shapes in a vast number of combinations. In addition, the animal for reasons of its own selects those objects and shapes which are of interest upon which to fix its gaze.

A cortical cell biased to some pattern (according to (2.2) and (2.3), for example) will, on the average, respond most strongly when its preferred pattern is in its receptive field. It seems reasonable to suppose that for a normal visual environment, the rest of the visual image is uncorrelated with the preferred pattern so that, averaged over the times the cortical cell is firing most strongly, *the visual image seen by the cortical cell, other than the preferred pattern might be regarded as noise.* (This now becomes a definition of a "normal" environment. Many deprivation experiments (cats raised with goggles, in planetaria, etc.) are designed precisely to alter this "normal" environmental situation.)

We therefore conclude that from the point of view of a cortical cell biased for one reason or another to respond preferentially to some specific pattern, the "normal" environment, excluding the preferred pattern, might be treated as noise; thus a visual pattern firing the cell, where the preferred e^k is one of the components, can be represented at a certain moment as

$$e^k(t) = e^k + \text{noise} \tag{2.4}$$

where noise is essentially unbiased and in the first approximation at least, can be assumed to be "white" or non-correlated from one time to another.

It now becomes meaningful to speak of K pattern classes, with the k^{th} class consisting of all the possible representations of pattern e^k with its noise-like visual accompaniment, i.e., of visual patterns like (2.4). In the mapping from retina to visual cortex, this pattern gives rise to a pair of parallel signal patterns $(d^k(t), b^k(t)) = (d^k + r(t), b^k + s(t))$. Since the noise components $r(t)$ and $s(t)$ are in fact representations of the random component $e^k(t) - e^k$, the best way to describe them is to use the mathematical correlation between $r(t)$ and $s(t)$, and assume that this correlation can have different values. On the other hand, since by definition $E(e^k(t)) = e^k$, or the random environment in $e^k(t)$ has no orientation bias, it is reasonable to assume that both $r(t)$ and $s(t)$ have zero means. (This is a matter of definition; if $r(t)$ and $s(t)$ do not have zero means, it suffices to redefine the noisy vs. patterned components in the external pattern $e^k(t)$ in such a way that, after the constant mapping of the visual pathway, the noise components have zero mean.)

The visual stimuli received by a dark-reared animal (or an animal deprived of patterned visual input by some procedure such as suturing the eyelids – though eyelid suturing and dark-rearing are not equivalent (Spear et al., 1978)) can probably be characterized as time-varying white noise. Such stimuli arrive on a cortical cell via modifiable or non-modifiable synaptic junctions so that we again denote them by $r(t)$ and $s(t)$. The signal patterns in this case are

$$(d^k(t), b^k(t)) = (r(t), s(t)) \tag{2.5}$$

The noise inputs, $r(t)$ and $s(t)$ may be mutually correlated at fixed times, but from one step, t, to another they both are zero-mean, independent, and at least weakly stationary. Also, the vector $r(t)$ will be assumed to have non-singular covariance structure.

III. Modifiability of Cortical Synapses

Synaptic modification dependent on inputs alone, of the type already directly observed in *Aplysia* (Kandel, 1976), is sufficient to construct a simple memory – one that distinguishes what has been seen from what has not, but does not easily separate one input from another (Anderson and Cooper, 1978). To distinguish between inputs as well requires synaptic modification dependent on information that exists at different places on the neuron membrane, what we call two (or higher) point modification. In order that such modification take place, information must be communicated from, for example, the axon hillock to the synaptic junction to be modified. This implies the possibility of internal communication of information within the neuron. One might guess that once the physiological mechanism for such communication was available, different types of two (or higher) point modification evolved in various ways. It is tempting to conjecture that a liberating evolutionary step was just the development of this means of internal communication which, coupled with the ability of synapses to modify, created the possibility for a new organization principle.

[2] The dependence of the inner product (b^i, b^l) or (d^i, d^l) with increasing angle has been discussed in Nass and Cooper, p. 14 (1975). It is seen that a rather sharp fall off is expected. In our simulations we have used such a set of inner products. However, the form of $Md + Zb$ at time zero is chosen to be more complicated than (2.1) to allow for broader initial tuning curves

A number of related articles on the development of cortical neurons have appeared recently (von der Malsburg, 1973; Nass and Cooper, 1975; Perez et al., 1975; Kohonen and Oja, 1976; Kohonen et al., 1977 and Anderson et al., 1977). In these, the modification hypothesis may be regarded as different realizations of the "conjunction rule" of Hebb (1949) or of the two-point modification mentioned above. Cooper (1973) and Nass and Cooper (1975) explored some of the consequences of a passive two-point modification of synaptic junctions. In Kohonen's models (1977) the concept of optimal mappings has been introduced. The error-correcting and noise-attenuating properties of the optimal associative mappings make such mappings better able to extract a basic pattern immersed in a noisy environment, and also to separate between the different pattern classes.

In this paper, we employ a variation of the two-point modification which captures some of the properties of passive modification (in that a cell can learn to respond or increases its response to a repeated external pattern) while at the same time modifying its response so that it responds to no more than one pattern. In this way, the tuning curve sharpens and the mapping from input to output becomes optimal.

Threshold Passive Modification

The general modification algorithm for the mapping, M may be written[3]

$$M(t)=\gamma M(t-1)+\delta M \tag{3.1}$$

In passive modification (Cooper, 1973) the change in strength of the labile junction M_{ij} of the efferent cell i is given by

$$\delta M_{ij}\sim(\text{output})_i(\text{input})_j \tag{3.2}$$

when the output is below a maximum (saturation) level, μ; there the output (excluding spontaneous fluctuations) is just the result of the input applied to already existing synapses. When the output is equal to or above the maximum level μ,

$$\delta M_{ij}=0 \tag{3.3}$$

so that above this maximum level, all of the synapses M_{i1}, M_{i2},... associated with the cell, i, stop modifying. (The only further change in synaptic strengths is then due to the uniform decay of all of the junctions if $\gamma<1$ (3.1).) This results in a maximum firing rate (which in any case is physiologically necessary) in such a way as to preserve the information in the synaptic junctions.

Applied to visual cortex, where the spontaneous firing rate is low (Herz et al., 1964), occasionally the incoming pattern is unable to make a particular cell fire. (The cell remains below threshold.) In such a case the modification of that cell's synapses is zero.

In what follows this passive modification algorithm is altered so that below a threshold called the modification threshold, θ_M, (which might, for example, be the actual threshold, θ_F, for the firing of the cell, the level of spontaneous activity, or some higher level) the synapses modify according to

$$\delta M_{ij}\sim-(\text{"output"})_i(\text{input})_j \tag{3.4}$$

Here the "output" is the actual output of the cell (on a scale where the mean spontaneous firing level is denoted by zero according to 1.3) if the cell is above the firing threshold, θ_F, or is just the integrated passive potential if the "output" is below the firing threshold θ_F. In this latter case we call this a force-no-fire (FNF) situation which results due to the integrated potential forcing the cell but not succeeding in firing it.

The modification threshold has the effect of increasing the response of a cell to an input to which it responds sufficiently strongly while decreasing its response (negative or positive) toward the level of spontaneous activity to inputs to which it responds too weakly. It is this which drives the system to an optimal mapping.

We thus assume that the modifiability of a synaptic junction is dependent on events that occur at different parts of the same cell and on the rate at which the cell responds: below θ_M, above θ_M but below the maximum rate μ, and above μ.

Suppose now that at time t the external pattern $e^k(t)$, $k\in(1,2,\dots K)$, appears and maps into the pattern pair $(d^k(t), b^k(t))$. Adding channel noise, including the effect of lateral inhibition[4] and using (1.4) we can write for the output of the i^{th} cell

$$c_i(t)=P[c_i^\kappa(t)] \tag{3.5a}$$

where

$$c_i^\kappa(t)=c_i^0(t)-\sum_{j\neq i}\kappa_{ij}c_j^0(t) \tag{3.5b}$$

and

$$c_i^0(t)=\sum_{j=1}^{L_d}M_{ij}(t-1)d_j^k(t)+\sum_{j=1}^{L_b}Z_{ij}b_j^k(t)+x(t) \tag{3.5c}$$

Here the κ_{ij} are the coefficients of lateral inhibition. It is convenient to write the modification algorithm from

[3] In this paper we use discrete time, $t=0,1,2\dots$ where one unit of time can be thought to correspond approximately to the duration of a burst of activity in the LGN and cortical cell axons. For further details, see NC

[4] We use here a form of "forward" lateral inhibition that leads to linear equations and is somewhat simpler than the form employed by NC. Our results are relatively independent of the precise form used

the point of view of this i^{th} cell. What we are primarily interested in then is the i^{th} row of the matrix M. Denoting this by m^i and the corresponding row of Z by z^i we can rewrite (3.5c) as

$$c_i^0(t)=(m^i(t-1),d^k(t))+(z^i,b^k(t))+x(t) \tag{3.5d}$$

With this the threshold passive modification algorithm becomes[1]

$$\begin{aligned}
&m^i(t)=\gamma m^i(t-1)+\eta^+ c_i^\kappa(t)d^k(t)\\
&\qquad \text{if} \quad \mu > c_i^\kappa(t) \geqq \theta_M;\\
&m^i(t)=\gamma m^i(t-1)\\
&\qquad \text{if} \quad c_i^\kappa(t) \geqq \mu;\\
&m^i(t)=\gamma m^i(t-1)-\eta^- c_i^\kappa(t)d^k(t)\\
&\qquad \text{if} \quad c_i^\kappa(t)<\theta_M
\end{aligned} \tag{3.6}$$

In this form the modification algorithm is discontinuous at μ. A more convenient form for analysis results if the change in synaptic modification at $c_i^\kappa=\mu$ does not take place abruptly, but rather weakens when output frequency approaches its maximum level. Therefore, the proportionality coefficient in front of (output) x (input) will be assumed to depend on output frequency in such a way that learning is fast when output is far from saturation, but tends to zero as output tends to saturation limit. A suitable functional relationship for this proportionality or gain factor would be

$$(\text{constant}) \times \left[\frac{\mu}{(\text{output})}-1\right] \tag{3.7}$$

which clearly decreases to zero as output grows to the saturation limit μ. As will be seen presently, the above functional form is especially suitable for mathematical treatment.[5] To show the relation between this modified algorithm and some algorithms previously studied (notably stochastic approximation type algorithms in regression problems and pattern recognition (Kohonen, 1977)) we substitute

$$\eta^+(t)=\eta^+\left(\frac{\mu}{c_i^\kappa(t)}-1\right)$$

for η^+ in (3.6). This now gives

$$\begin{aligned}
&m^i(t)=\gamma m^i(t-1)+\eta^+(\mu-c_i^\kappa(t))d^k(t)\\
&\qquad \text{if} \quad \mu > c_i^\kappa(t) \geqq \theta_M\\
&m^i(t)=\gamma m^i(t-1)\\
&\qquad \text{if} \quad c_i^\kappa(t) \geqq \mu\\
&m^i(t)=\gamma m^i(t-1)-\eta^- c_i^\kappa(t)d^k(t)\\
&\qquad \text{if} \quad c_i^\kappa(t)<\theta_M
\end{aligned} \tag{3.8}$$

Now, if threshold is surpassed, an *additive* term μ appears in (3.8).

This makes the algorithm considerably easier to handle from a mathematical point of view.

IV. Acquisition and Loss of Specificity

An Ideal Case: Input of Patterns Without Noise

To show as clearly as possible how threshold passive modification can lead to an optimal mapping for which each cell responds only to a single orientation pattern and thus becomes specific (Imbert and Buisseret, 1975; Blakemore and Van Sluyters, 1975; Buisseret and Imbert, 1976; Frégnac and Imbert, 1977, 1978) or "sharply tuned" we first consider an ideal case. In this the cortical neurons receive only stimuli due to pure patterned inputs – the fixed pattern pairs (d^k, b^k) without noise and without specific inclusion of effects of other cells in the same orientation column (excitation) or cells in nearly orientation columns (lateral inhibition). Such other effects are discussed later.

As a starting-point of the analytical approach, assume that in the initial situation ($t=0$) the leading pattern pair (d^1, b^1) produces an above threshold response, $c_1 > \theta_M$, in neuron one while the response of this first neuron to the rest of the pattern pairs remains below modification threshold:

$$\begin{aligned}
&(m^1(0),d^1)+(z,b^1)>\theta_M\\
&(m^1(0),d^j)+(z,b^j)<\theta_M \qquad j \neq 1\,.
\end{aligned} \tag{4.1}$$

In what follows $m(t)$ and z are understood to mean $m^1(t)$ and z^1 respectively.

The actual stimuli received by a cortical cell contain at least two distinct "noise-like" components: one due to fluctuations in the firing rates of the neurons, variations in spontaneous activity, the other due to the visual background in which the optimal pattern (e.g., a bar or an edge) is immersed. The addition of such noise components leads to responses above θ_M for patterns other than the optimal one (d^1, b^1).

For $\theta_M=\theta_F$ (the force-no-fire situation) noise is required to make the cell fire at all for non-optimal input so that in the ideal case the cell is in a sense already "sharply tuned" since it does not respond to any pattern other than the leading pattern pair (d^1, b^1). If θ_M is at the level of spontaneous activity, noise is required to give response levels above the level of spontaneous activity for non-optimal input. If θ_M is above the level of spontaneous activity one would obtain such responses to non-optimal inputs in the absence of noise. (Experimental observation of the variance in the level of spontaneous activity (Bienenstock and Frégnac, Private Communication)

[5] A somewhat more complicated form makes the algorithm continuous at θ_M as well as μ without substantially altering our results

indicates that normal variations are large enough to produce responses to non-optimal input. In our simulations we have used noise levels consistent with such observed variance.)

Equation (3.8) can be regarded as yielding a Markov-process with continuous states and discrete time, where the probability measures are in principle derivable from the statistical properties of the input process (the fixed pattern pairs enter in a random order) and the initial state. No mean-square convergence or convergence in probability of $m(t)$ to a fixed value is possible for the algorithm of (3.8). There will be fluctuations which in the real world might be reflected in a certain percentage of neurons failing to achieve sharp tuning and remaining aspecific or losing their ability to respond entirely. It should be stressed that while, for simplicity, only one of the modifiable cells is considered here, the overall effect of orientation specificity is due to the entire network of neurons.

To obtain asymptotic values at large times analytically, we concentrate our attention on the "successful" neurons. A linearization of the algorithm then offers a fairly good approximation; this is confirmed by simulations. Here linearization means the linearization of the recursion for $E[m(t)]$, the expectation of $m(t)$. While in reality the choice of strengthening vs. weakening of the response, according to the threshold modification hypothesis, is statistically dependent on $m(t)$ itself, we now make the approximation that this choice depends only on the input. This leads to a linear algorithm for $E[m(t)]$. Otherwise the ensueing nonlinearities prevent us from obtaining any closed-form recursive relationship from which the asymptotics might be derived with reasonable ease.

Theorem 1. *Let*

$$m(t+1)=\gamma m(t)+\eta^{+}[\mu-(m(t),d^1)-(z,b^1)]d^1$$

if the pattern pair (d^1,b^1) *enters* (4.2)

and

$$m(t+1)=\gamma m(t)-\eta^{-}[(m(t),d^j)+(z,b^j)]d^j$$

if the pattern pair $(d^j,b^j), j\neq 1$, *enters.* (4.3)

Also let $\eta^{+}>0$, $\eta^{-}>0$, $1\geqq\gamma>0$. *Let* d^i*-vectors be linearly independent and let each pair* (d^i,b^i) *appear with equal probability at step* t, *independent of the input at previous step* $t-1$.

Then if η^{+} *and* η^{-} *are small enough, the vector* $\sigma(t)$ *with elements*

$$E[(m(t),d^k)]+(z,b^k)\qquad(k=1,2,\ldots,K),$$

giving the average responses of the cell to pattern pairs (d^k,b^k), *tends asymptotically to*

$$\bar{\sigma}=-[(1-\gamma)I+H]^{-1}(1-\gamma)y+\mu c^1,\tag{4.4}$$

where H *is a square matrix with*

$$H_{ij}=\frac{\eta^{-}}{K}(d^i,d^j)\qquad(j>1),$$

$$H_{i1}=\frac{\eta^{+}}{K}(d^i,d^1),\tag{4.5}$$

y *is a* K*-dimensional vector with*

$$y_j=-(z,b^j)\qquad(j>1),$$

$$y_1=\mu-(z,b^1),\tag{4.6}$$

and c^1 *is the unit vector* $(1\ 0\ 0\ldots0)^T$.

The proof of this theorem is given in the Appendix.

The meaning of the asymptotic result is not easily visualized, if $\gamma<1$. However, since the limits of the average responses to the K different pattern classes are continuous functions of γ as $\gamma\to 1$, a good approximation for γ close to one (which is the realistic situation) is given by the vector μc^1, whose elements are μ, 0, 0...0. Thus the response of the cell remains on average at saturation, μ, for the leading pattern pair, but tends to zero (or, in terms of actual firing frequencies, to the level of spontaneous activity) for all the other pattern pairs. This is an optimal state. It should be observed that if there are patterns (d^j,b^j) that are geometrically close to the leading pair (d^1,b^1), the response for them will nevertheless tend to zero which means that only a very narrow part of the visual environment manages to elicit responses that deviate from the spontaneous activity level. So the tuning is indeed sharp.

To obtain a better picture of the asymptotic result of (4.4) when γ is close to but not quite equal to one, a simple corollary to Theorem 1 is given in the following.

Corollary 1. *In Theorem* 1, *let* $1-\gamma=\varepsilon$ *be small. Then in* (4.4)

$$\bar{\sigma}=\mu c^1-\varepsilon H^{-1}y+0(\varepsilon^2),\tag{4.7}$$

where H, y, *and* c^1 *are defined in Theorem* 1.

The proof is given in the Appendix.

In the simulation of Figs. 6a and b, the initial situation was such that the response of the cell under study was slightly above the modification threshold for one of the pattern pairs (say d^1,b^1) and below this threshold for the others.

The modifiable part of the memory mapping was now changed in the simulation according to (3.8). The simulation shows that the response to the leading pattern pair (d^1,b^1), to which the neuron was sensitive from the very start, increases towards saturation level causing strong firing; at the same time, the other responses decrease toward the level of spontaneous activity. This sharpening results in an optimal state in

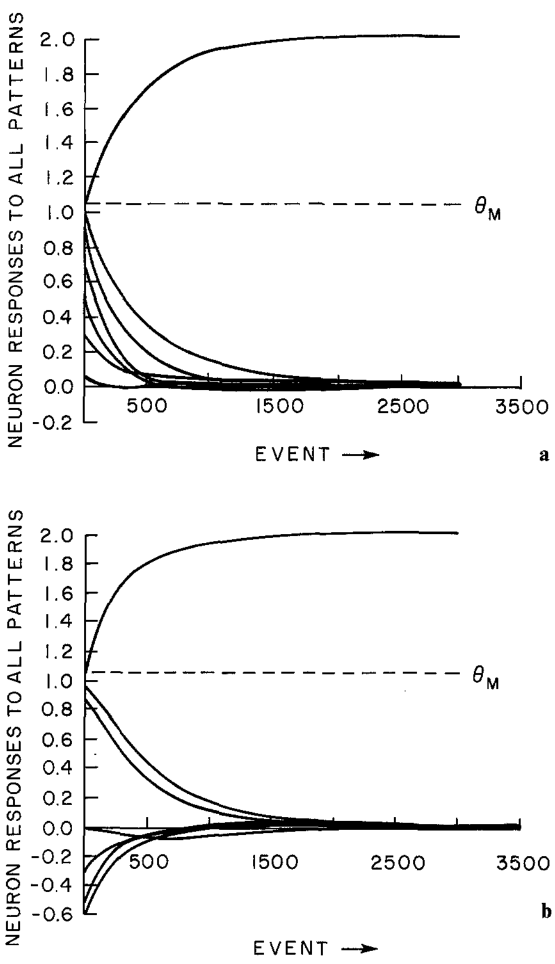

Fig. 6a and b. The responses of a neuron to 7 different noiseless patterns as functions of time. In these simulations, the parameters have the following values: $\gamma=1.0$, $\eta^+=0.032$, $\eta^-=0.017$, $\mu=2.0$, $\theta_M=1.05$, and $\kappa=0$. All noise levels were zero. The individual patterns entered in a pseudorandom order such that within each set of 7 steps in the algorithm, each pattern enters once but otherwise in a random sequence. z is such that it alone gives response 1 to the leading pattern and 0.5 to all the other patterns. Upper curves: the response to the leading pattern to which the initial response was higher than θ_M. Lower curves: responses to the other 6 patterns. In **a** all of the initial responses of the cell to the input patterns were positive; in **b** some of the initial responses of the cell were negative. The only difference between **a** and **b** is in the value of $m(0)$

which the leading response is close to saturation while the other responses are close to zero. Because of the finite length of simulation runs, the exact asymptotic values and especially their dependence on the system parameters is not always evident; however, they are given by Theorem 1 above.

In all of our simulations the vectors d were set equal to the vectors b. The dependence of the inner products of these vectors (b^i, b^l) or (d^i, d^l) with increasing angle was chosen to duplicate the sharp fall-off discussed in NC and assumes no directional preference.[2] The inner products used were $(b^i,b^l)=(d^i,d^l)=f(i-l)$; $f(0)=1$, $f(1)=0.4$, $f(2)=0.3$, $f(3)=0.2$, $f(4)=0.2$, $f(5)=0.3$, $f(6)=0.4$.

In Fig. 7, the synaptic strengths were "frozen" at given times between inputs, showing what the postsynaptic potentials of the cell would be if the noiseless ideal patterns were input. In the simulation there are 7 cells and 7 input classes. These classes might be thought of as modeling single vectors produced by an illuminated bar on the retina, with 7 different orientations at 15° intervals, as explained in Sect. II. The cell shown is typical in its response. The patterns to the right of the leading one correspond to increments of 15° in the angle of the bar; the response of the cell to those to the left, corresponding to decrements of 15°, has been set symmetric to those on the right. The 13 points then show the postsynaptic potentials produced by the pure patterns, i.e., the function

$$(m,d^k)+(z,b^k)$$

as k is varied but m and z are the fixed "frozen" values at that step. These points have been connected by straight lines.

Input of Patterns with Noise

With the inclusion of noise the consequences of threshold passive modification are more difficult to analyze. There are several problems. It is hard to estimate the relative strengths of the actual external stimulus $(d(t), b(t))$ and channel noise $x(t)$. It is also difficult to estimate in the external stimulus the relative strengths of the fixed standard inputs (d^i, b^i) and the rest of the image field $(r(t),\ s(t))$, reflecting the changing background in which the primary patterns always appear in normal visual experience. A safe method seems to be to choose the noise components to be relatively large; if the system works in a satisfactory way with a great deal of noise, it is likely to give even better results in less noisy situations.

Even when there is an initial tendency for a cell to respond more strongly to a single pattern class than to the others, noise of large magnitude can cause many misfirings and the cells might pick up more than one pattern class, especially if the inner products between different patterns d^i and d^j are large. A similar situation was analyzed by Nass and Cooper (1975). There the introduction of lateral inhibition between cortical cells limited the number of cells that respond to a single pattern, and the combination of lateral inhibition with the upper limit beyond which no modification occurs limits the number of patterns to which a single cell responds. We employ a similar lateral inhibition here. Thus the innate tendency for a single cell to respond most strongly to a single pattern is enhanced by lateral inhibition between cells in different orientation columns. At the same time it is expected that excitation within an orientation column will enhance the response of a cell in that column to the "preferred" pattern.

In our simulations, in spite of the fact that there are a considerable number of misfirings the neurons very seldom become asymptotically sensitive to other pattern classes than their own. The simulations show that the lateral inhibition term (which in any case was small) has a considerable effect on modification only in the beginning, tending to disappear rapidly as learning continues. Then the contribution of lateral inhibition is small compared to the threshold passive modification effect, which alone is sufficient to drive the neurons towards a limiting state that is a distorted version of the optimal mapping obtained in the noiseless case.

Let $m^i(t)$ and z^i be the modifiable and non-modifiable synaptic vectors of the i^{th} cell at time t. Then including lateral inhibition the response before thresholding, proportional to the post-synaptic potential of the cell, is

$$\begin{aligned} c_i^{\kappa}(t+1) = {} & [(m^i(t), d(t+1)) \\ & +(z^i, b(t+1)) + x_i(t+1)] \\ & - \sum_{j \neq i} \kappa_{ij} [(m^j(t), d(t+1)) \\ & +(z^j, b(t+1)) + x_j(t+1)] \end{aligned} \tag{4.8}$$

where $(d(t+1), b(t+1))$ is the pair of input vectors, $x_i(t+1)$ is the channel noise input to the i^{th} cell and κ_{ij} are lateral inhibition coefficients.

We now write again

$$d(t+1) = d^k + r(t+1)$$

$$b(t+1) = b^k + s(t+1)$$

where k denotes the pattern class input at $(t+1)^{\text{st}}$ step. Neglecting the term due to lateral inhibition we have for $c_i^{\kappa}(t+1)$

$$\begin{aligned} & (m^i(t), d^k) + (z^i, b^k) + (m^i(t), r(t+1)) \\ & \quad +(z^i, s(t+1)) + x_i(t+1). \end{aligned} \tag{4.9}$$

There $(m^i(t), d^k) + (z^i, b^k)$ would be the post-synaptic potential if no noise were present, and the three other components in the sum represent noise.

In case all elements of $r(t+1)$ and $s(t+1)$ and $x_i(t+1)$ are non-correlated and zero-mean we would obtain for the variance of the noise component

$$\begin{aligned} \operatorname{var}(\text{noise}) = {} & \sum_{j=1}^{L_d} m_j^i(t)^2 \operatorname{var}(r_j(t+1)) \\ & + \sum_{j=1}^{L_b} (z_j^i)^2 \operatorname{var}(s_j(t+1)) \\ & + \operatorname{var}(x(t+1)). \end{aligned} \tag{4.10}$$

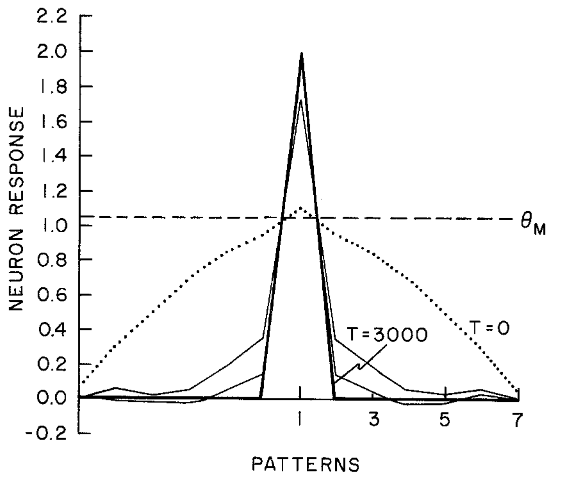

Fig. 7. The responses of a neuron to 7 different noiseless patterns at intervals from $t=0$ to $t=3000$. (This is the same simulation as that shown in Fig. 6a.) The responses of the neuron to patterns on the left side of pattern 1 were set equal to the responses to the corresponding patterns on the right. This give a symmetrical tuning curve. (In general we would obtain this result in an actual simulation. However, machine time available did not permit running simulations with thirteen patterns.)

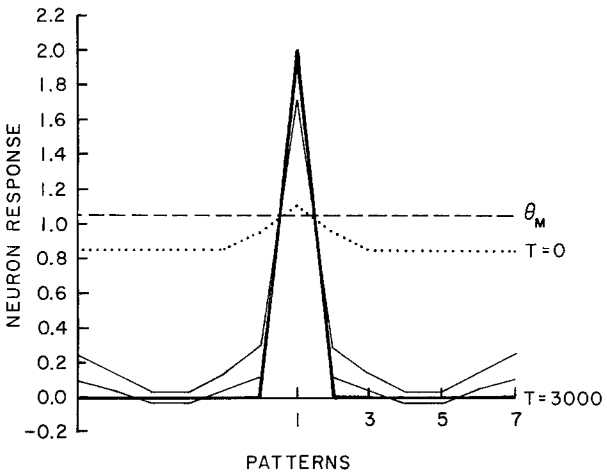

Fig. 8. The responses of a neuron to seven different noiseless patterns. This simulation is identical to that of figure seven except that $m(0)$ is chosen to make the initial responses of the neuron relatively flat (as is the case for some aspecific cells)

Furthermore, if $\operatorname{var}(r_j(t+1)) = V_1 = \text{constant}$

$$\operatorname{var}(s_j(t+1)) = V_2 = \text{constant}$$

and

$$\operatorname{var}(x(t+1)) = V_3 = \text{constant}$$

we obtain

$$\operatorname{var}(\text{noise}) = \|m^i(t)\|^2 V_1 + \|z^i\|^2 V_2 + V_3. \tag{4.11}$$

If there is correlation between $r(t)$ and $s(t)$, then (4.11) will not hold exactly. To take a simple special case, assume that the cross-covariance of r and s is diagonal, i.e., $\operatorname{cov}[r_i(t+1), s_j(t+1)] = \delta_{ij} V_4$, with δ_{ij} the Kroenecker delta. Then it is easy to show that we have

$$\operatorname{var}(\text{noise}) = \|m^i(t)\|^2 V_1 + \|z^i\|^2 V_2 + V_3 + 2(m^i(t), z^i) V_4 \tag{4.12}$$

So the passive potential on the cell can be divided into two parts: the noiseless part due to pure input patterns, and a noise component with zero mean and variance given by the above equations. Since noise is the sum of several independent components, its distribution is close to the normal distribution.

The dependence of the asymptotic limits on the different parameters can again be approximated by a mathematical result of a linearized algorithm, where the starting point is a sub-optimal situation with low probabilities of misfirings. A modification of Theorem 1 leads to the following result:

Theorem 2. *Let $m(t)$ be the vector of junction strengths of one of the cells, say, the one for which (d^1, b^1) are the leading patterns. Let $m(t)$ change according to*

$$m(t) = \gamma m(t-1) + \eta^+ [\mu - (m(t-1), d^1(t)) - (z, b^1(t)) - x(t)] d^1(t)$$

if $(d^1(t), b^1(t))$ enters;

$$m(t) = \gamma m(t-1) + \eta^- [-(m(t-1), d^i(t)) - (z, b^i(t)) - x(t)] d^i(t)$$

if $(d^i(t), b^i(t))(i \neq 1)$ enters. Let the parameters satisfy $0 < \gamma \leqq 1$, $\eta^+ > 0$, $\eta^- > 0$. Assume the following for the inputs: for each i, $d^i(t) = d^i + r(t)$, $b^i(t) = b^i + s(t)$, where $E[r(t)] = 0$, $E[s(t)] = 0$, $E[r(t) \times r(t)] = R$ is positive definite and bounded, $E[s(t) \times s(t)]$ is bounded, and the probability of occurrence at step t of a pattern from the i^{th} class is uniform, i.e., $\frac{1}{K}$, and independent of the previous step. Also, assume that channel noise $x(t)$ satisfies $E[x(t)] = 0$ and is independent of $r(t)$ and $s(t)$. Let d^i be linearly independent and let $m(0)$ have finite mean.

Then for η^- and η^+ sufficiently small

$$\lim_{t \to \infty} E(m(t)) = U^{-1} a, \tag{4.13}$$

where U is the matrix

$$U = R + d^1 \times d^1 + \frac{\eta^-}{\eta^+} \sum_{j=2}^{K} (R + d^j \times d^j) + \frac{K}{\eta^+}(1-\gamma) I, \tag{4.14}$$

and a is the vector

$$a = [\mu - (z, b^1)] d^1 - p - \frac{\eta^-}{\eta^+} \sum_{j=2}^{K} [(z, b^j) d^j + p], \tag{4.15}$$

with

$$p = E[r(t) \times s(t)] z.$$

Proof. See Appendix.

By the above the limit of $m(t)$ in mean sense is close to the optimal vector; it becomes optimal, as can be expected, if γ tends to one (no forgetting) and the noise in d-patterns is zero, as is shown by Theorem 1 and Corollary 1. In the general case, noise tends to distort even the mean value from the ideal case obtained with noiseless learning and given by Theorem 1. However, the sharpening of the tuning curve is evident. Small residual responses of a cell to non-optimal patterns are likely related to spontaneous activity.

In the following simulations, the maximal intensity of the channel noise $x(t)$ at cortical cell inputs has been approximately one half of the input; this is consistent with experimental data (Bienenstock and Frégnac, Private Communication). If smaller channel noise is used the convergence is better. (In an experimental situation the visual inputs are usually adjusted to include an optimal input and to be relatively free of variations. The reported standard deviations then give a rough picture of the fluctuations of the firing frequency in response to noise-less input patterns, i.e., what has been termed here as channel noise.)

The relative intensities of the patterns vs. background were varied; the maximum background-to-pattern ratios $\|r(t)\|/\|d^j\|$ and $\|s(t)\|/\|b^j\|$ were well over one. When this situation is considered geometrically in the L_d- and L_b-dimensional pattern spaces, it becomes clear that this kind of "noise" is already very high. Assume, for example, that the inner product of two unit length pattern vectors d^i and d^j is 0.4, which is a typical value and corresponds to an angle of 66.5° in the pattern space. (This is not the same as the angle between the corresponding visual images e^i and e^j on the retina, due to the complicated way in which the two-dimensional visual patterns are mapped to the frequency – coded signal vectors d^i and d^j.) If noise vectors with larger than unit length and completely random directions are now added to the patterns d^i and d^j, a very considerable overlapping of the pattern regions takes place. In fact d^j could even be considered as a noisy version of d^i and vice versa.

In Fig. 9 simulation results in the noisy case are shown. This plot was constructed along the same principles as Fig. 7. The noiseless patterns (d^k, b^k) have been used in computing the plot so that a comparison with Fig. 7 would be possible; however, when m^i has been modified in this run, then the noisy patterns have been used. At some points in Fig. 9, the noise limits have been plotted. Each individual realization of a "tuning curve" would be inside these limits.

Input of Noise Without Patterns

A picture of the functioning of threshold modification now emerges: the neurons are driven towards the optimal feature extraction state by patterned input information immersed in a noise-like sensory environment and carried by noisy neural pathways. The final state is one in which there is sharp tuning; each cell is very sensitive to one pattern class and almost completely insensitive to the other classes. The post-synaptic potential caused by the leading pattern is large, leading to almost maximal firing frequency, while the post-synaptic potentials to other patterns tend to be such that even large variations do not cause significantly intense firings.

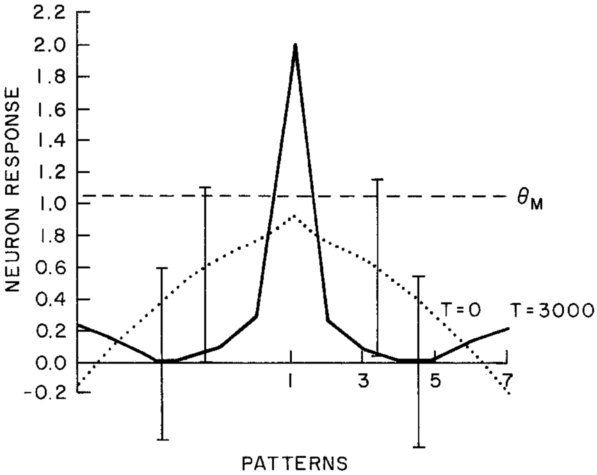

Fig. 9. The responses of a neuron to the 7 patterns at times $t=0$ and $t=3000$. In this simulation, the input vectors were the patterns with noise. Channel noise was also present. All elements of noise vectors $r(t)$ and $s(t)$ were uncorrelated and uniformly distributed over $[-0.3, +0.3]$. The input vectors used were of the form $d^k+r(t)$, $b^k+s(t)$. Channel noise $x(t)$ was uncorrelated with signal noise and uniformly distributed over $[-0.5, +0.5]$. Different input classes $(k=1,2,\ldots,7)$ entered in a pseudorandom order. Dotted curve: responses at time $t=0$. Solid curve: responses at time $t=3000$. The vertical bars represent regions, inside which the responses would be when channel noise is added. Therefore each individual realization of a "tuning curve" would be inside these regions, with the curves giving the average tuning at the respective times $t=0$ or $t=3000$. The parameters used in this simulation were: $\gamma=0.9999$, $\eta^+=0.035$, $\eta^-=0.017$, $\kappa=0.3$, $\mu=2$, $\theta_M=1.05$

We now consider what happens if the flow of patterned information ceases. There is some loss of information due to the factor γ. However, the experimental situation suggests that the effective γ should be very close to one (if not equal to one) since cortical cells are known experimentally to continue to respond to visual stimuli even if the animal is dark reared. There is another effect, however, which follows from threshold passive modification that drives the cells from sharp to broad tuning. The asymptotic state, however, is not simply zero but is more subtle, depending on correlations among the noise inputs.

If the flow of normal patterned visual stimuli ceases (for example due to dark-rearing or eyelid suturing) input to cortical cells continues but is changed in nature. When the eyelids are sutured diffuse light falls on the retina and produces occasional firings of the retinal ganglion cells. When the animal is dark-reared there still occur dark discharges of retinal cells. The resulting stimuli, however, are totally different from those assumed previously: instead of certain fixed patterns in a noise-like environment, the input now can be considered to be totally noise-like. We can denote the signal patterns caused by these noise-like visual stimuli in the two fixed parallel pathways as $r(t)$ and $s(t)$. There is also channel noise present emanating from the signal-carrying pathway itself, although the intensity of this channel noise may be different from that connected with strong visual stimuli. Let channel noise be, as before, $x(t)$.

Once more omitting the lateral inhibition term the adaptation algorithm for one of the cells is, according to (3.8)

$$\begin{aligned}
m(t) &= \gamma m(t-1) + \eta^+ [\mu - (m(t-1), r(t)) \\
&\qquad - (z, s(t)) - x(t)] r(t) \\
\text{if} \quad & \mu > (m(t-1), r(t)) + (z, s(t)) + x(t) > \theta_M \\
m(t) &= \gamma m(t-1) \\
\text{if} \quad & (m(t-1), r(t)) + (z, s(t)) + x(t) > \mu \\
m(t) &= \gamma m(t-1) - \eta^- [(m(t-1), r(t)) \\
&\qquad + (z, s(t)) + x(t)] r(t) \\
\text{if} \quad & (m(t-1), r(t)) + (z, s(t)) + x(t) < \theta_M
\end{aligned} \tag{4.16}$$

To get an analytical picture of the situation, the behavior is again approximated by a linear algorithm. Assume in the above equation that during one step of the recursion, corresponding to an integration period of about 1 second, the noisy input is changing so rapidly around its mean value zero compared to the rate of change of synaptic strengths, that the integrated or averaged input activity alone is not sufficient to fire the cell with a frequency higher than the modification threshold and cause synaptic growth. So in effect we concentrate on the third of (4.16), omitting the two low-probability cases.

The following theorem allows a variety of values for the possible correlation of $x(t)$ with $r(t)$ and $s(t)$, as well as the correlation between the two inputs $r(t)$ and

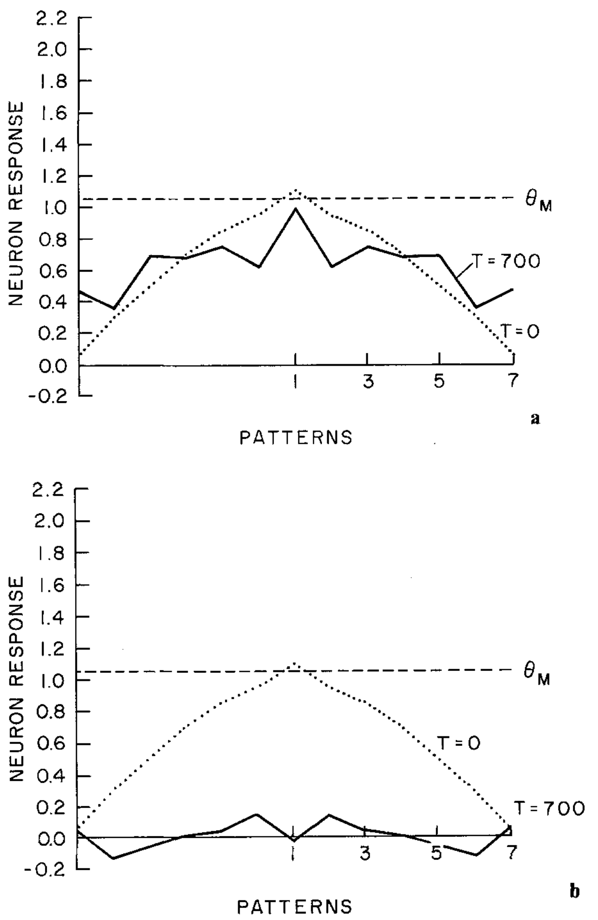

Fig. 10a and b. The response of a neuron to signal and channel noise. The channel noise was uncorrelated with the signal noise and was uniformly distributed over $[-0.5, +0.5]$. The signal noise ($r(t)$, $s(t)$) was uniformly distributed over $[-0.3, +0.3]$. In **a** the two signal noise inputs, $r(t)$ and $s(t)$ were uncorrelated, while in **b** they were completely correlated ($r(t)=s(t)$). The parameters in this simulation were: $\gamma=1$, $\eta^+=0.035$, $\eta^-=0.5$, $\mu=2$, $\theta_M=1.05$, $\kappa=0$. η^- was increased in order to decrease simulation time. Dotted upper curve: responses at time$=0$. Solid curve: responses at time$=700$

$s(t)$. It is shown there how the asymptotic mean value of vector $m(t)$ is related to the various parameters and statistical correlations, and how choosing the gain factor η^- small improves the asymptotic mean square error.

Theorem 3. *Consider the algorithm*

$$m(t)=\gamma m(t-1)-\eta^-[(m(t-1),r(t)) + (z,s(t))+x(t)]r(t) \qquad (4.17)$$

There $r(t)$ and $s(t)$ are sequences of independent random vectors and $x(t)$ is a sequence of independent random scalars satisfying: $E[r(t)]\equiv 0$, $E[s(t)]\equiv 0$, $E[r(t)\times r(t)]\equiv R$ is positive definite and $r(t)$, $s(t)$, and $x(t)$ are bounded for each t. Let $E[r(t)\times s(t)]z\equiv p$, and $E[x(t)\times r(t)]\equiv q$.

Let $m(0)$ have finite mean and covariance and let $0<\gamma\leqq 1$.

Then there exists a number $\delta>0$ such that if $0<\eta^-<\delta$,

$$\lim_{t\to\infty} E[m(t)]=\bar{m}=-\eta^-[(1-\gamma)I+\eta^- R]^{-1}(p+q) \qquad (4.18)$$

and

$$E[\|m(t)-\bar{m}\|^2]<\omega \qquad (4.19)$$

for all t with ω a finite constant. Furthermore, $\limsup_{t\to\infty} E[\|m(t)-\bar{m}\|^2]$ as a function of η^- tends to zero as $\eta^-\to 0$.

Proof. See Appendix.

Several conclusions can be drawn from the asymptotic result given in the theorem. If it is desirable to have the limit as close to zero as possible, this would imply that the gain should be small, the cross-correlations of the random processes weak and forgetting strong, as is of course clear from physical considerations. On the other hand, strong cross-correlations and large values of the two system parameters lead to a "noisy" limit with mean value away from zero, possibly closer to the negative of the non-modifiable part, with the effect that after a period with only noise stimulating the cells the "tuning" tends to be even less sharp than that given by the nonplastic part of the mapping only.

Figure 10a shows a simulation run using the original algorithm in (4.16). $s(t)$, $r(t)$ and $x(t)$ had zero mean. There was no correlation between noise values at different times.

Although no sensory patterns were now used as input, the plots in Figs. 10a and b were constructed along the same principles as in ones in Figs. 7 and 9. The plot at a given step reveals what the response of the cell would be if the values were again "frozen" and patterned input were used to determine how the cell is responding. It is seen that the sharpening, shown in Figs. 7 and 9 and obtained using patterned input, is now practically reversed in Fig. 10a; the tuning curve seems to settle to a wide tuning with the non-modifiable z-part of the mapping predominant. Simulations showed that now there will be firings occasionally, but they are not selective and do not manage to maintain the tuning.

In Fig. 10b, another simulation of the non-patterned situation is shown. The difference from the simulation of Fig. 10a is that in b there is correlation between $r(t)$ and $s(t)$; to emphasize the effect $r(t)$ and $s(t)$ were set equal for each t. Theorem 3 predicts that the total response of the cell will tend on the average to zero, i.e., $m(t)$ tends now to $-z$ in the mean sense. This is what in fact happens: however η^- had to be chosen to be large in this run so that the fluctuations in $m(t)$ are rather large.

Reversibility of Gain and Loss of Specificity

We now ask whether a return of patterned stimuli after a period of darkness would result in regaining specificity of response to these stimuli (see the experimental result of Fig. 4). This is indeed the case in the present model.

The simulation in Figs. 11 and 12 exhibit a complete run, where a model neuron is shown to gain sharp tuning with noiseless patterned input, then lose it again under noise-like input, but regain it when the noiseless patterned stimuli are once more allowed to appear in the input. For the first 3000 time units (the actual length of this period depends on the values of various parameters) the neuron receives input consisting of patterns without noise. A sharpening takes place. For the next 500 steps the patterns are absent in the input and only uncorrelated noise is presented. Partial loss of specificity occurs. Then, for another period of 2500 steps, noiseless patterns appear again and the modification is seen to be very similar to the original sharpening effect in the beginning of the run. This is what might be expected, since the modification scheme and the inputs are the same and there are only small differences in starting values in these two sharpening periods in the plot.

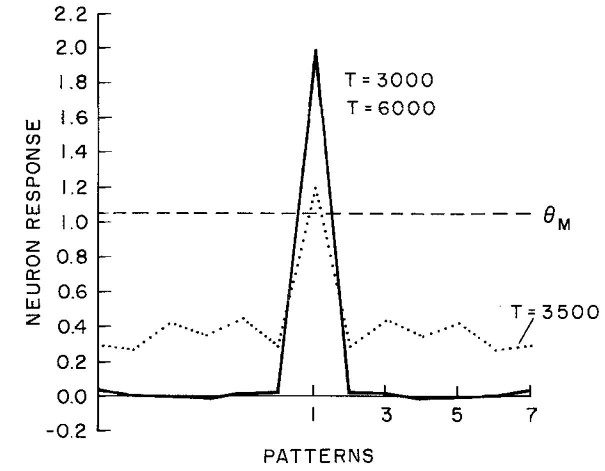

Fig. 11. Loss and retrieval of specificity. At $T=3000$ the neuron is sharply tuned (Heavy line). Noise was presented from $T=3001$ to $T=3500$. The channel noise was uniformly distributed over $[-0.5, +0.5]$. The signal noise was uniformly distributed over $[-0.3, +0.3]$ and was uncorrelated. Parameters used were, $\gamma=1$, $\eta^+=0.035$, $\eta^-=0.1$, $\theta_M=1.05$, $\mu=2$, $\kappa=0$. At $T=3500$ the cell was broadly tuned (Dotted line). Then (noiseless) patterns were presented from $T=3501$ to $T=6000$; η^- was changed back to 0.017 to make the result comparable with Fig. 7. At $T=6000$ (Solid line) the neuron has regained all of its previous sharp tuning. [The parameter η^- is varied to decrease simulation time. For the same η^- loss of specificity is slower than increase of specificity.]

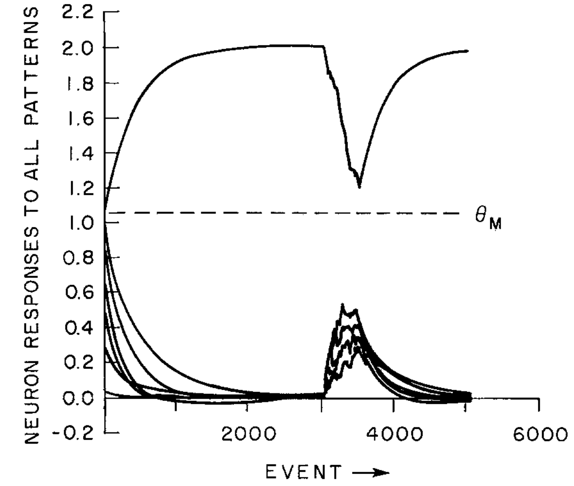

Fig. 12. Loss and retrieval of specificity. This figure is plotted in a manner similar to Fig. 6a and b. The parameters used are those of Fig. 11. From $T=0$ to $T=3000$ noiseless patterns were presented and the neuron acquired a sharp tuning. Uncorrelated noise was presented to the neuron from $T=3001$ to $T=3500$ and a broadening (i.e., loss of specificity) occurred. Noiseless patterns were presented from $T=3501$ to $T=6000$ and the neuron regained its former sharp tuning (i.e., retrieval of specificity)

V. Conclusion and Discussion

We have assumed that between lateral geniculate and visual cortical cells there exist labile synapses which are modified according to the threshold form of passive modification as well as non-labile synapses which contain permanent information. These latter give the cortical cells a weak initial tendency to fire more strongly and readily to some orientations in the visual field than to others.

In the theory that results, there is an increase in the specificity of the response of a cortical cell to visual input (sharpening of its tuning curve) when that cell has been exposed to stimuli that are the results of normal patterned visual experience. When exposed to noise-like input, such as might be expected when an animal is dark-reared or raised with eyelids sutured, there is a loss of specificity. Specificity can be regained, however, with a return of input due to patterned vision. This seems to us to provide a possible explanation of the experimental results obtained by Imbert and Buisseret (1975); Blakemore and Van Sluyters (1975); Buisseret and Imbert (1976); and Frégnac and Imbert (1977, 1978).

In addition to this basic behaviour, simulations and mathematical results on the asymptotic states of the neural network show some more subtle phenomena that depend upon values of system parameters, notably the amount of decay (forgetting per unit time), the strength of selective modification of synaptic junctions, and the different statistical properties of noise factors. In the following discussion, some of these effects are illuminated by making simplifying approximations.

Sharpening of Tuning with Learning

The theoretical outcome of a long period of learning with patterned stimuli is given in Theorems 1 and 2. Sharpest tuning is achieved if forgetting is very slow, i.e., if memory traces tend to be rather permanent over short periods at least, and if the patterns occurring in the course of learning are as noiseless representations of a set of prototypes as possible. Then the modifiable part, M, of the memory mappings tends to become such that it practically cancels out the effect of the non-modifiable part Z, except for the optimal stimuli which give a sharp and high peak in the tuning curve. The response to these optimal patterns is close to the maximum firing level.

With stimuli corresponding to normal visual input, the modification of the tuning curve of a cell is monotone, going from broadly tuned to sharply tuned. The neuron changes from a non-specific unit, showing some response to all input patterns, to an immature one, for which some orientations have already lost the ability to elicit any kind of response above the spontaneous activity of the cell. Then the range of orientations for which the neuron is sensitive diminishes further until only a high and narrow peak, centered at the leading pattern for that cell, is visible in the tuning curve: the neuron has become a specific unit. This is in good agreement with experimental observation of the progression from a-specific to immature to specific as reported by Frégnac and Imbert (1977, 1978). In addition, this suggests that the classifications (Imbert and Buisseret, 1975; Buisseret and Imbert, 1976) of a-specific, immature and specific are relatively arbitrary divisions in what is really a continuum of response.

In learning with patterned stimuli, the outcome is not very sensitive to system parameters; the gross behavior of sharpening of tuning close to the optimum was achieved in simulations under a considerable range of parameter values. It is especially notable that even with relatively high noise levels, with signal-to-noise ratios considerably smaller than one, we still obtain qualitatively very similar behaviour to the noiseless learning cases. This confirms the good averaging properties of the modification algorithm.

Loss of Specificity

With noise-like stimuli, corresponding to non-patterned visual input, widening of the cortical cell tuning curve takes place, practically independent of how sharp the tuning was originally. The main effect is a rapid decay of the central peak of the tuning curve; the change in response for non-optimal patterns is smaller. The overall time constant of the loss of specificity was larger than that of gain of specificity.

This seems to be in agreement with experimental results, where giving patterned stimuli to dark-reared animals produces remarkably fast gains of specificity (see Fig. 4), while the loss of specificity is a more gradual process.

In this situation the effect of parameters is more important. In the absence of the overwhelming influence of strong patterned stimuli, some rather subtle phenomena, due to correlation properties among different noise components, now become visible. To illustrate these effects the following simplifications are made in Theorem 3:

We let the two parallel noise-like input vectors, $r(t)$ and $s(t)$, have equal numbers of elements ($L_a = L_b$) and for all i let these elements satisfy

$$\begin{aligned} &\operatorname{var}(r_i(t)) = \operatorname{var}(s_i(t)) = V_1\,, \\ &\operatorname{cov}(r_i(t), s_i(t)) = V_2\,, \end{aligned} \tag{5.1}$$

but all the other correlations between elements are zero. This might correspond to a structural situation in which each modifiable synapse on the cortical neuron has its non-modifiable counterpart, and the inputs $r_i(t)$ and $s_i(t)$ entering these two synapses are correlated. In fact if $V_1 = V_2$ in (5.1), then this correlation is complete (essentially $r_i(t) = s_i(t)$) and we see a situation where these two synapses, one modifiable and the other non-modifiable, share the same input signal.

We denote possible correlation between channel noise x and input r_i (for all i) by

$$\operatorname{cov}(x(t), r_i(t)) = Q\,.$$

Then Theorem 3 gives the limit of $E(m(t))$ as

$$\bar{m} = \frac{-\eta^-}{(1-\gamma) + \eta^- V_1}(V_2 z + q) \tag{5.2}$$

where q is a vector with each element equal to Q.

Equation (5.2) has several implications. If $1-\gamma$ is very small and Q can be neglected, we have approximately

$$\bar{m} \approx -\frac{V_2}{V_1} z\,. \tag{5.3}$$

This shows that strong correlation ($V_2 \approx V_1$) leads to a limit $\bar{m}$ such that the *total* response of the cell will be close to zero, while weak correlation ($V_2 \approx 0$) drives $E(m(t))$ close to zero, leaving the non-modifiable part z alone to determine the tuning of the cell. If, however, $1-\gamma$ is not small and Q is not negligible, then $\bar{m}$ may be close to zero even in the correlated case.

We see then that if there is strong correlation between r and s, the noise inputs to labile and non-labile junctions, the initial bias of the cell can be entirely lost. A return of specificity would then not necessarily be to the same orientation as that preferred originally. (If excitation is assumed between cells in an orientation column, other cells which retained their original bias might still guide a cell that has lost its bias to its original orientation preference. However, an entire column could shift its orientation preference (Blakemore and Van Sluyters, 1974; Movshon, 1976)).

It should be stressed that the theorems give only weak convergence, so $\bar{m}$ in (5.2) would be only an average limit in a large number of similar cells. Although with η^- small the actual limits tend to be close to $\bar{m}$ according to Theorem 3, the non-zero variance in $\bar{m}$ would cause variations in asymptotic tuning in individual cells. In some cases our model neuron might even change its preferred pattern. It is very important which pattern happens to give the highest response when learning with patterned input is again commenced, since it is this initial situation alone which determines the optimal pattern for the given cell. It should also be mentioned that we have not taken into account effects due to binocular interactions. Such effects are at present being investigated and will be discussed in a future publication.

If the theory presented here bears any relation to the facts, then the interesting relation of the amount of loss of specificity to correlations between different noisy inputs might provide information on the connectivity of the neural network as well as providing a variety of opportunities for further experimental verification.

VI. Appendix

Proof of Theorems 1 and 2

Since the algorithm in Theorem 1 is just a special case of that in Theorem 2 (when $s(t)$, $r(t)$ and $x(t)$ have zero variance), the proofs can be combined. Take the algorithm of Theorem 2:

$$\begin{aligned} m(t) = \gamma m(t-1) + \eta^+ [\mu - (m(t-1), d^1(t)) \\ - (z, b^1(t)) - x(t)] d^1(t) \end{aligned}$$

if $(d^1(t), b^1(t))$ enters;

$$\begin{aligned} m(t) = \gamma m(t-1) + \eta^- [-(m(t-1), d^i(t)) \\ - (z, b^i(t)) - x(t)] d^i(t) \end{aligned}$$

if $(d^i(t), b^i(t))$ $(i \neq 1)$ enters.

Calculate first the conditional expectation value

$$E\{[(m(t-1), d^i(t)) + (z, b^i(t)) + x(t)]\, d^i(t) | m(t-1)\}$$

using the assumptions made on $b^i(t)$ and $d^i(t)$. We obtain

$$\begin{aligned} E\{[(m(t-1), d^i) + (m(t-1), r(t)) + (z, b^i) \\ + (z, s(t)) + x(t)]\, (d^i + r(t)) | m(t-1)\} \\ = (d^i \times d^i) m(t-1) + R m(t-1) + (z, b^i) d^i + p\,. \end{aligned}$$

When these are multiplied by the class probabilities $\frac{1}{K}$ and summed, noting the difference when $i=1$, we obtain

$$\begin{aligned} E[m(t)|m(t-1)] = \frac{1}{K} \gamma m(t-1) \\ + \frac{1}{K} \eta^+ [\mu d^1 - (d^1 \times d^1) m(t-1) - R m(t-1) \\ - (z, b^1) d^1 - p] \\ + \sum_{j=2}^{K} \left\{ \frac{1}{K} \gamma m(t-1) - \frac{1}{K} \eta^- [(d^j \times d^j) m(t-1)) \right. \\ \left. + R m(t-1) + (z, b^j) d^j + p] \right\}, \end{aligned}$$

yielding

$$\begin{aligned} E[m(t)|m(t-1)] = \gamma m(t-1) + \\ \frac{\eta^+}{K} \cdot [\mu d^1 - (d^1 \times d^1) m(t-1) \\ - R m(t-1) - (z, b^1) d^1 - p] \\ - \frac{\eta^-}{K} \sum_{j=2}^{K} [(d^j \times d^j) m(t-1) \\ + R m(t-1) + (z, b^j) d^j + p] \\ = \bar{U} m(t-1) + \bar{a}\,, \end{aligned}$$

with

$$\begin{aligned} \bar{U} = \gamma I - \frac{\eta^+}{K} [R + (d^1 \times d^1)] \\ - \frac{\eta^-}{K} \sum_{j=2}^{K} [R + (d^j \times d^j)] \end{aligned} \qquad (A.1)$$

and

$$\begin{aligned} \bar{a} = \frac{\eta^+}{K} [\mu d^1 - (z, b^1) d^1 - p] \\ - \frac{\eta^-}{K} \sum_{j=2}^{K} [(z, b^j) d^j + p]\,. \end{aligned} \qquad (A.2)$$

Taking expectations over $m(t-1)$ yields $E(m(t)) = \bar{U} E(m(t-1)) + \bar{a}$. Since $\gamma \leqq 1$, $\eta^+ > 0$, $\eta^- > 0$, R positive definite and finite, it is obvious that the norm of $\bar{U}$ will be smaller than one if η^+ and η^- are not too large. Then the equation is stable and the solution tends to the fixed point $\lim_{t \to \infty} E(m(t)) = (I - \bar{U})^{-1} \bar{a}$, which is seen to be equal to $U^{-1} a$ when both $I - \bar{U}$ and $\bar{a}$ are multiplied with $\frac{K}{\eta^+}$.

To derive from this Theorem 1, set $R=0$, $p=0$ in $\bar{U}$ and $\bar{a}$. Then $\lim_{t\to\infty} E(m(t))=\bar{m}$ is a solution of the equation

$$(I-\bar{U})\bar{m}=\bar{a}$$

or

$$\left[I-\gamma I+\frac{\eta^+}{K}(d^1\times d^1)+\frac{\eta^-}{K}\sum_{j=2}^{K}(d^j\times d^j)\right]\bar{m}$$

$$=\left[\frac{\eta^+}{K}(\mu-(z,b^1))d^1-\frac{\eta^-}{K}\sum_{j=2}^{K}(z,b^j)d^j\right].$$

That $\bar{m}$ indeed satisfies (4.4) of Theorem 1, is seen if we denote by D the matrix with columns $d^1, d^2, \ldots, d^K$, by Λ the diagonal matrix with diagonal elements $\frac{\eta^+}{K}, \frac{\eta^-}{K}, \frac{\eta^-}{K}, \ldots, \frac{\eta^-}{K}$, and by y the vector with elements $\mu-(z,b^1)$, $-(z,b^2), \ldots, -(z,b^K)$. With this notation the above equation reads $[(1-\gamma)I+D\Lambda D^T]\bar{m}=D\Lambda y$. Multiplying by D^T from the left yields $[(1-\gamma)I+D^TD\Lambda]D^T\bar{m}=D^TD\Lambda y$. This in turn implies

$$[(1-\gamma)I+D^TD\Lambda]D^T\bar{m}=[(1-\gamma)I+D^TD\Lambda]y-(1-\gamma)y,$$

yielding

$$D^T\bar{m}-y=-[(1-\gamma)I+D^TD\Lambda]^{-1}(1-\gamma)y,$$

$$=-[(1-\gamma)I+H]^{-1}(1-\gamma)y, \qquad \text{(A.3)}$$

where $H=D^TD\Lambda$ is easily seen to have elements H_{ij} given by (4.5) in Theorem 1.

The nonsingularity of matrix $(1-\gamma)I+H$ follows from the fact that matrix $(1-\gamma)D^TD+D^TD\Lambda D^TD$ $=[(1-\gamma)I+H]D^TD$ is non-singular because of the matrix inversion lemma (see Kohonen, 1977). Since also D^TD is nonsingular because D has linearly independent columns, then $(1-\gamma)I+H$ must be nonsingular.

Since the first element of vector $D^T\bar{m}-y$ is $\lim_{t\to\infty} E[(m(t),d^1)]+(z,b^1)-\mu$ (from (4.6)), and the i^{th} element ($i>1$) is $\lim_{t\to\infty} E[(m(t),d^i)]+(z,b^i)$, we see that

$$\bar{\sigma}=D^T\bar{m}-y+\mu c^1. \qquad \text{(A.4)}$$

Substituting this in (A.3) establishes Theorem 1.

Proof of Corollary 1

With $1-\gamma=\varepsilon$ small, (4.4) yields

$$\bar{\sigma}=-(\varepsilon I+H)^{-1}\varepsilon y+\mu c^1$$
$$=-\varepsilon[(\varepsilon H^{-1}+I)H]^{-1}y+\mu c^1$$
$$=-\varepsilon H^{-1}(I+\varepsilon H^{-1})^{-1}y+\mu c^1$$
$$=-\varepsilon H^{-1}(I-\varepsilon H^{-1}+\varepsilon^2H^{-2}-\ldots)y+\mu c^1$$
$$=-\varepsilon H^{-1}y+0(\varepsilon^2)+\mu c^1,$$

as was to be shown.

Proof of Theorem 3

Denote $v(t)=m(t)-\bar{m}$, where $\bar{m}$ is the vector $-\eta^-[(1-\gamma)I+\eta^-R]^{-1}(p+q)$. Then $\bar{m}$ satisfies $\bar{m}=\gamma\bar{m}-\eta^-R\bar{m}-\eta^-(p+q)$. Substituting $v(t)$ in the recursion yields $v(t)=m(t)-\bar{m}=\gamma m(t-1)-\gamma\bar{m}$

$$-\eta^-[(m(t-1)-\bar{m},r(t))+(\bar{m},r(t))+(z,s(t))$$
$$+x(t)]r(t)+\eta^-[R\bar{m}+p+q]$$
$$=\gamma v(t-1)-\eta^-[(r(t)\times r(t))v(t-1)$$
$$+(r(t)\times r(t))\bar{m}-R\bar{m}$$
$$+(r(t)\times s(t))z-p+x(t)r(t)-q].$$

Now taking expectations over $r(t)$, $s(t)$, $x(t)$, but keeping $v(t-1)$ fixed yields

$$E[v(t)|v(t-1)]=\gamma v(t-1)-\eta^-[Rv(t-1)$$
$$+R\bar{m}-R\bar{m}+p-p+q-q]=(\gamma I-\eta^-R)v(t-1),$$

since $v(t-1)$ depends only on $s(k)$, $r(k)$, $x(k)$ for $k\leqq t-1$, hence is independent of $r(t)$.

Finally taking expectations over $v(t-1)$ yields

$$E[v(t)]=(\gamma I-\eta^-R)E[v(t-1)].$$

Now, $E[v(t)]$ converges to zero if and only if $\|\gamma I-\eta^-R\|<1$. This is true if $\gamma-\eta^-\lambda_0<1$ and $\gamma-\eta^-\lambda_1>-1$, where λ_0 and λ_1 are the smallest and largest eigenvalues of R. Since $\gamma\leqq1$, $\lambda_0>0$, it is sufficient for the first inequality that $\eta^->0$; for the second inequality it is sufficient that $\eta^-<\frac{1+\gamma}{\lambda_1}$ which is positive since λ_1 is finite, by the assumption on boundedness of $r(t)$.

To show uniform boundedness of $E[\|m(t)-\bar{m}\|^2]=E[\|v(t)\|^2]$ we write

$$\|v(t)\|^2=\gamma^2\|v(t-1)\|^2$$
$$+(\eta^-)^2\|u\|^2-2\gamma\eta^-(u,v(t-1)),$$

where

$$u=(r(t)\times r(t))v(t-1)+(r(t)\times r(t))\bar{m}$$
$$-R\bar{m}+(r(t)\times s(t))z-p+x(t)r(t)-q.$$

Now it would be possible to limit the variance explicitly. However, since this limit would not be a tight one anyway, let us use a briefer method: the form of u and the boundedness of $r(t)$, $s(t)$ and $x(t)$ reveal that there exist finite non-negative constants ω_0, ω_1, ω_2 such that

$$E[\|u\|^2|v(t-1)]\leqq\omega_0\|v(t-1)\|^2$$
$$+\omega_1\|v(t-1)\|+\omega_2.$$

The expectation of $(u, v(t-1))$ is easier to handle: we obtain

$$
\begin{aligned}
&E[(u,v(t-1))|v(t-1)]\\
&\quad=(v(t-1), Rv(t-1))+(v(t-1), R\bar{m})\\
&\qquad-(v(t-1), R\bar{m})+(v(t-1), p)\\
&\qquad-(v(t-1), p)+(v(t-1), q)\\
&\qquad-(v(t-1), q)\geqq\|v(t-1)\|^2\lambda_0 .
\end{aligned}
$$

Combining these yields

$$
\begin{aligned}
&E[\|v(t)\|^2|v(t-1)]\leqq\gamma^2\|v(t-1)\|^2\\
&\quad+(\eta^-)^2[\omega_0\|v(t-1)\|^2\\
&\quad+\omega_1\|v(t-1)\|+\omega_2]-2\gamma(\eta^-)\lambda_0\|v(t-1)\|^2\\
&\quad=[\gamma^2+(\eta^-)^2\omega_0-2\gamma\lambda_0\eta^-]\|v(t-1)\|^2\\
&\quad+(\eta^-)^2[\omega_1\|v(t-1)\|+\omega_2].
\end{aligned}
$$

It is easily established that $(\eta^-)^2\omega_0-2\gamma\lambda_0\eta^-+\gamma^2-1=0$ has two real roots in η^-, one non-positive, one positive, between which the expression is negative. Also, because of continuity, there is a number $\delta>0$ such that $(\eta^-)^2\omega_0-2\gamma\lambda_0\eta^-+\gamma^2\in(0,1)$ as long as $0<\eta^-<\delta$. Let us pick η^- from this interval and denote then

$$\alpha=(\eta^-)^2\omega_0-2\gamma(\eta^-)\lambda_0+\gamma^2, \quad 0<\alpha<1 .$$

Now taking expectations with respect to $v(t-1)$ yields

$$
\begin{aligned}
&E[\|v(t)\|^2]\leqq\alpha E[\|v(t-1)\|^2]\\
&\quad+\omega_1(\eta^-)^2E[\|v(t-1)\|]+\omega_2(\eta^-)^2 .
\end{aligned}
$$

By Jensen's inequality,

$$E[\|v(t-1)\|]\leqq(E[\|v(t-1)\|^2])^{1/2} .$$

The boundedness of $E[\|v(t)\|^2]$ follows from the fact that

$$
\begin{aligned}
&\alpha E[\|v(t-1)\|^2]+\omega_1(\eta^-)^2\\
&\quad\cdot(E[\|v(t-1)\|^2])^{1/2}+\omega_2(\eta^-)^2
\end{aligned}
$$

with

$$\alpha\in(0,1), \eta^->0, \omega_1>0, \omega_2>0 ,$$

becomes smaller than $E[\|v(t-1)\|^2]$ when the latter is large enough, and in this region $E[\|v(t)\|^2]<E[\|v(t-1)\|^2]$. So it is clear that $E[\|v(t)\|^2]$ will always be bounded by some number ω.

Finally, consider the recursion for $E[\|v(t)\|^2]$ as η^- tends to zero. Since ω cannot increase as η^- decreases, we can write

$$
\begin{aligned}
&E[\|v(t)\|^2]\leqq\alpha E[\|v(t-1)\|^2]+\omega_1(\eta^-)^2\omega^{1/2}\\
&\quad+\omega_2(\eta^-)^2=\alpha E[\|v(t-1)\|^2]+(\eta^-)^2\omega_3
\end{aligned}
$$

for all η^- small enough. There $\omega_3=\omega_1\omega^{1/2}+\omega_2$ is a non-negative constant. The solution satisfies

$$
\begin{aligned}
&E[\|v(t)\|^2]\leqq\alpha^tE[\|v(0)\|^2]\\
&\quad+(\eta^-)^2\omega_3\sum_{i=1}^{t}\alpha^{t-i}=\alpha^tE[\|v(0)\|^2]\\
&\quad+(\eta^-)^2\omega_3\frac{1-\alpha^t}{1-\alpha},
\end{aligned}
$$

tending to $\frac{(\eta^-)^2\omega_3}{1-\alpha}$ as $t\to\infty$ for all sufficiently small $\eta^->0$.

Letting $\eta^-\to 0$ establishes the last part of the theorem, since

$$\frac{(\eta^-)^2\omega_3}{1-\alpha}=\frac{(\eta^-)^2\omega_3}{1-\gamma^2+2\gamma\lambda_0\eta^--\omega_0(\eta^-)^2}\to 0 \text{ as } \eta^-\to 0$$

(for $\gamma=1$, notice that $\lambda_0>0$, R being positive definite).

Acknowledgements. We would like to thank our colleagues in the Center for Neural Science for their encouragement and advice. In particular, we thank Mr. Elie Bienenstock for carefully reading the manuscript and suggesting several improvements. One of the authors (LNC) would like to express his appreciation to the Fondation de France and the CNRS for their support during his stays in France. He would also like to express his appreciation to his colleagues in the laboratory of Michel Imbert at the College de France for their hospitality and for the patience with which they explained the intricacies of the experimental results. In addition, financial support to one of the authors (EO) from the Academy of Finland and the Emil Aaltonen Foundation is gratefully acknowledged.

References

Anderson, J.A.: Two models for memory organization using interacting traces. Math. Biosci. **8**, 137–160 (1970)

Anderson, J.A.: A simple neural network generating an interactive memory. Math. Biosci. **14**, 197–220 (1972)

Anderson, J.A., Silverstein, J.W., Ritz, S.A., Jones, R.S.: Distinctive features, categorical perception, and probability learning. Some applications of a neural model. Psychoanal. Rev. **84**, 413–451 (1977)

Anderson, J.A., Cooper, L.N.: Les modeles mathematiques de l'organization biologique de la memoire. Plurisci. 168–175 (1978)

Batini, C., Buisseret, P.: Sensory peripheral pathway from extrinsic eye muscles. Arch. Ital. Biol. **112**, 18–32 (1974)

Bienenstock, E., Frégnac, Y.: Stability of response of single cells in kittens visual cortex. (to be published)

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

Blakemore, C., Mitchell, D.E.: Environmental modification of the visual cortex and the neural basis of learning and memory. Nature **241**, 467–468 (1973)

Blakemore, C., Van Sluyters, R.C.: Reversal of the physiological effects of monocular deprivation in kittens. Further evidence for a sensitive period. J. Physiol. (London) **237**, 195–216 (1974)

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

Buisseret, P., Imbert, M.: Visual cortical cells. Their developmental properties in normal and dark reared kittens. J. Physiol. (London) **255**, 511–525 (1976)

Buisseret, P., Gary-Bobo, E., Imbert, M.: Ocular motility and recovery of orientational properties of visual cortical neurons in dark-reared kittens. Nature **272**, 816–817 (1978)

Cooper, L.N.: A possible organization of animal memory and learning. In: Proceedings of the Nobel Symposium on Collective Properties of Physical Systems. Lundquist, B., Lundquist, S., eds. London, New York **24**, 252–264 (1973)

Frégnac, Y., Imbert, M.: Cinetique de developement des cellules du cortex visuel. J. Physiol. (Paris) **6**, T.73 (1977)

Frégnac, Y., Imbert, M.: Early development of visual cortical cells in normal and dark-reared kittens. Relationship between orientation selectivity and ocular dominance. J. Physiol. (London) **278**, 27–44 (1978)

Frégnac, Y.: Cinetique de development du cortex visuel primaire chez le chat. Effets de la privation visuelle binoculaire et modele de maturation de la selective a l'orientation. Doctoral thesis, Université René Descartes (1978)

Hebb, D.O.: The organization of behavior. New York: Wiley 1949

Henry, G.H., Dreher, B., Bishop, P.O.: Orientation specificity of cells in cat striate cortex. J. Neurophysiol. **137**, 1394–1409 (1974)

Herz, A., Creutzfeldt, O., Fuster, J.: Statistische Eigenschaften der Neuronaktivität im ascendierenden visuellen System. Kybernetik **2**, 61–71 (1964)

Hirsch, H.V.B., Spinelli, D.N.: Modification of the distribution of receptive field orientation in cats by selective visual exposure during development. Exp. Brain Res. **12**, 509–527 (1971)

Hubel, D.H., Wiesel, T.N.: Receptive fields of single neurons in the cat striate cortex. J. Physiol. (London) **148**, 574–591 (1959)

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

Imbert, M., Buisseret, P.: Receptive field characteristics and plastic properties of visual cortical cells in kittens reared with or without visual experience. Exp. Brain Res. **22**, 2–36 (1975)

Kandel, E.R.: Cellular basis of behavior. San Francisco: Freeman 1976

Kohonen, T.: Correlation matrix memories. IEEE Trans. Comput. C-**21**, 353–359 (1972)

Kohonen, T.: Associative memory – a system – theoretical approach. Berlin, Heidelberg, New York: Springer 1977

Kohonen, T., Oja, E.: Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron like elements. Biol. Cybernetics **21**, 85–95 (1976)

Kohonen, T., Lethiö, P., Rovamo, J., Hyvärinen, J., Bry, K., Vainio, L.: A principle of neural associative memory. Neuroscience **2**, 1065–1076 (1977)

Movshon, J.A.: Reversal of the physiological effects of monocular deprivation in the kittens visual cortex. J. Physiol. (London) **261**, 125–174 (1976)

Nass, M., Cooper, L.N.: A theory for the development of feature detecting cells in visual cortex. Biol. Cybernetics **19**, 1–18 (1975)

Perez, R., Glass, L., Shlaer, R.J.: Development of specificity in the cat visual cortex. J. Math. Biol. **1**, 275–288 (1975)

Pettigrew, J.D., Freeman, R.D.: Visual experience without lines. Effects on developing cortical neurons. Science **182**, 599–601 (1973)

Pettigrew, J.D.: The effect of visual experience on the development of stimulus specificity by kitten cortical neurons. J. Physiol. **237**, 49–74 (1974)

Spear, P.D., Tong, L., Langsetmo, A.: Striate cortex neurons of binocularly deprived kittens respond to visual stimuli through the closed eyelids. Brain Res. **155**, 141–146 (1978)

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

Received: December 1, 1978

Prof. Dr. L. N. Cooper
Center for Neural Science
Brown University
Providence, R.I. 02912, USA