CHAPTER 24

State-Dependent Factors Influencing Neural Plasticity: A Partial Account of the Critical Period

P. W. MUNRO

PLASTICITY AND LEARNING

In this chapter I discuss the state dependence of learning rate using the development of ocular dominance in visual cortex (Area 17) as a model system. Several qualities make this system an excellent example of neural plasticity. Foremost and most obvious of these qualities is the clear and abundant evidence of plasticity in Area 17. Secondly, the susceptibility of ocular dominance properties to a very broad range of manipulations has been extensively analyzed and the circuitry connecting the retina to visual cortex is better understood than any other neural system of comparable complexity. Furthermore, for the purposes of this analysis, the system can be characterized in terms of two-dimensional "ocularity states," which serve to illustrate many of the mathematical concepts in this book with a relatively simple representation.

An account of early learning is presented in this chapter that can be described at several different levels. While the underlying concepts are actually quite simple, they are also in their essence mathematical and are thus expressed in mathematical language. To help make the presentation more understandable to all readers, I have tried to elaborate most of the points in more intuitive terms.

The aim of this chapter is to suggest that the high plasticity frequently observed in the early stages of a learning process may reflect, at least in part, an expected reduction in the modifiability of neuronal

response characteristics as their synaptic connections are strengthened as a result of experience. This contrasts with previously proposed alternative possibilities that this phenomenon is either maturational or that it is centrally gated by global modulatory signals.

The organization of the chapter is as follows: The present section introduces the reader to the subject and the approach of the chapter. The following section describes a general developmental phenomenon called the *critical period* during which neural networks are observed to be relatively plastic, then focuses on how plasticity is manifested during the development of ocular dominance in visual cortex, both during and after the critical period. In the third section, the reader is introduced to the notion of state-dependent plasticity, a concept that is then applied to give an account of the time course of ocular dominance plasticity. In the fourth section, these ideas are compared with experimental data. In the final section, the ramifications of state-dependent plasticity are discussed in relation to experimental data, existing theories, and their predictive power.

Focus on single-unit learning. Unlike most of the chapters in this book, which explore the behavior of whole networks, the present chapter is focused at the level of the single neuron and its afferent synapses. This does not reflect a difference in belief about the appropriate level of analysis, but rather it is grounded in the simple pragmatic observation that at the single-unit level, a fairly good account of the relevant phenomena can be given. Put another way, the critical period is not seen as an emergent property of an interactive network in this approach, but as a phenomenon that might be traceable to the effects of prior experience on the behavior of individual units.

Learning rate as a factor in PDP models. Notions of learning rate (in general) and the critical period (in particular) bear on models for learning by parallel distributed processing systems and have not been extensively studied by theorists apart from specific applications. A popular approach to selecting a learning rate is the pragmatic strategy in which one chooses a step size that yields a good solution in an acceptable time frame; however, rate properties of adaptive systems deserve more attention, as they may provide an important link between theoretical models and empirical data. Several existing theories of plasticity in cortex are concerned with the sequence of learning and ignore the *rate* of learning, either ignoring the critical period or making ad hoc assumptions to account for it. Conversely, in and of themselves, the scope of the ideas to be presented in this chapter is limited to rate and do not include an explicit rule for synaptic modification. Since it lacks such an explicit learning rule (that could be used to predict final states), it must

be emphasized that the present *partial account* is not sufficiently complete to constitute either a theory or a model of cortical plasticity.

Two cautionary notes regarding terminology. First, it should be noted that although the terms *plasticity* and *learning* are used interchangeably in this chapter, there is a difference in nuance between them. Plasticity carries with it a sense of physical modifiability but not necessarily toward adaptation, whereas learning has a flavor of storing information from the environment with less emphasis on the neural (synaptic) substrate. Both, however, refer to *adaptation*, an environment-driven metamorphosis that enables a system to generate responses appropriate to that environment. Second, one should be particularly careful in interpreting the term *state* (especially as in "state of the system") here. In many of the network models, the state describes the instantaneous activation levels of all the units. Here, the same term refers to the connectivity values of the afferents to a unit. In the context of a network model, the *activation state* is generally changing very rapidly as the assembly of units settles in response to a particular stimulus according to information implicit in the *connectivity state*, which modifies more slowly under the influence of some self-organizing adaptive principle.

THE CRITICAL PERIOD

It is well established that plasticity varies over the course of ontogenetic development. During so-called *critical* or *sensitive* periods of plasticity, certain perceptual[1] mechanisms have been found to be more susceptible to adaptive modification by the stimulus environment, whereas the susceptibility is low or nonexistent at other times. Sensitive periods have been demonstrated in many systems, such as song learning in birds (Konishi, 1978), socialization in dogs (Scott, 1962, 1968), and phoneme discrimination in humans (Eimas, 1978; Miyuwaki et al., 1975). However, the phenomenon has been most extensively analyzed with respect to the development of binocular interaction in mammalian (particularly cat and monkey) visual cortex.

It is certainly not expected that all of these examples will necessarily have the same physiological substrate, nor would it be correct to

[1] Weak evidence exists for sensitive periods at "higher" (i.e., more cognitive) levels of mental function. The most compelling results are based on findings that children recover from trauma-induced aphasia more rapidly and completely than adults (who rarely, if ever, fully recover). For an overview, see Lenneberg (1967).

assume that any behavioral observation of sensitive periods in learning should be reflected by sensitive periods in neurophysiological measures of single-unit plasticity. However the converse inference, namely, that plasticity in single units is sufficient (but not necessary) for macroscopic learning phenomena, seems valid. Therefore the reader should keep in mind that the framework to be presented below for neuronal plasticity is not restricted to visual cortex, but might be applicable much more generally.

Ocular Dominance and its Susceptibility to the Environment

Interest in visual system plasticity and the generation of relevant data enjoy a symbiotic relationship that has led to the production of a great deal of both. A brief overview is given in this section of a part of this data that is based on measuring changes in the *ocular dominance* (OD) statistics of a large sample of neurons (from about 50 to over 1,000, depending on the experiment). This method, devised by Wiesel and Hubel (1963), consists of assigning each unit to one of seven ocular dominance classes based on the relative values of its optimal responses to monocular inputs from both eyes. Since there is an innate statistical bias for a neuron to respond preferentially to the eye on the opposite side of the head, the ocular dominance classes are ordered according to the influence on the neuron of the ipsilateral eye relative to the contralateral eye. A Class 1 neuron responds exclusively to input from the eye contralateral to it, a Class 7 neuron responds only to the ipsilateral eye, a Class 4 neuron gives roughly equal responses to each eye, and neurons in the other classes are responsive to both eyes with corresponding degrees of preference.

The stimulus environment can be crudely controlled by suturing shut neither, either, or both of the eyelids. The resulting paradigms are respectively called normal rearing (NR), monocular deprivation (MD), and binocular deprivation (BD). These can be combined into rearing schedules such as reverse suture (RS) in which a period of MD is followed by opening the closed eye and suturing shut the previously open eye, and alternating monocular occlusion (AM) which consists of several such reversals.

Several results can be interpreted in terms of an age-defined critical period, though the boundaries of the critical period seem to vary somewhat with the particular details of the experimental procedure. Hubel and Wiesel (1970b) found that MD drives all neurons to strongly prefer the open eye, provided the animal is still sufficiently young—neuronal sensitivity to MD seems to peak at about 4 weeks and disappears by about 3 months (see Figure 1). More recently, other researchers

FIGURE 1. The ODH shift associated with MD. The classic result of Hubel and Wiesel is shown in this figure adapted from their 1970 paper. The ODH is shown for a kitten deprived of vision in one eye from Day 10 to Day 37. Units that cannot be driven through either eye are classified as visually unresponsive (VU) and are represented here by an unfilled bar. (The ODH of a normally reared [NR] kitten is nearly symmetric.)

(Cynader, Timney, & Mitchell, 1980) report an ocular dominance shift in kittens subjected to monocular deprivation for 3 months after rearing under normal conditions from birth to 5 months of age. Because it is such a striking effect, requiring only 24 hours or less during the height of the critical period (Olson & Freeman, 1975; Peck & Blakemore, 1975), the ocular dominance shift under monocular deprivation is often used as a test for assessing cortical plasticity.

With the exception of BD, other rearing schedules distort the relatively uniform *ocular dominance histogram* (ODH) of normal kittens and cats. Binocular deprivation generally leads to a histogram with all OD classes represented, as in newborn and normally reared cats. However, unlike normally reared cats, newborn and BD cats have neuronal responses described as "sluggish" and "unpredictable" with ill-defined receptive fields (Cynader, Berman, & Hein, 1976; Wiesel & Hubel, 1965). In another rearing paradigm, artificial strabismus (AS), the eyes receive noncorrelated inputs. Both surgical (severing some of the extraocular muscles) and optical (attaching eyeglasses with prisms) techniques have been used to simulate strabismus. Strabismic kittens exhibit U-shaped ocular dominance histograms, i.e., the neurons become highly monocular, but do not show an overall preference for one eye over the other. While AS yields a sensitivity time course that closely matches MD (Hubel & Wiesel, 1965), the same cannot be said of all procedures.

The results reviewed thus far appear consistent with a fairly strict maturational hypothesis for cortical sensitivity. However, evidence against this view comes from a very interesting result that has been obtained by Cynader et al. (1976) who have raised binocularly deprived

kittens throughout and beyond the period associated with plasticity (up to 10 months) and then opened just one eye. They observed a strong shift in ocular dominance after a 3-month period of monocular experience. Cynader (1983) extended this result by raising the kittens in the dark for up to 2 years, followed by 9 to 18 months of monocular experience. This phenomenon, described at the end of this section, provides the most compelling evidence for a link between a neuron's connectivity state and the plasticity of the state.

Plasticity-Modulating Processes

The hypothesis that plasticity is strictly age-related falls down in the face of another kind of evidence as well; it has been found that plasticity can be modulated by a variety of chemical and/or physical processes (see Table 1). One might suppose that the critical period is explicitly controlled by one or more of these processes (i.e., it is part of the "design" of the system). However, as will be seen, this is not the only alternative hypothesis. For now, consider the following evidence concerning the roles of a variety of modulators of plasticity, focusing on norepinephrine in particular.

Norepinephrine. Catecholamines attracted attention as potential modulators of neuronal plasticity during the 1960s (see, for example, Crow, 1968). The catecholamine norepinephrine (NE) is pervasively distributed throughout the central nervous system from a localized region in the brain stem, the *locus coeruleus* (LC); thus the anatomy of the norepinephrine system makes it an attractive candidate for a global

TABLE 1

PROCESSES THAT MODULATE CORTICAL PLASTICITY

Natural Site of Modulation	Experimental Stimulus Modality	References
Locus Coeruleus	pharmacological	Kasamatsu & Pettigrew (1979)
Proprioception (Extraocular)	mechanical	Freeman & Bonds (1979) Buisseret et al. (1978)
Reticular Formation	electrical	Singer (1979)
Medial thalamic nuclei	electrical	Singer (1982) Singer & Rauschecker (1982)

modulator of plasticity. The LC/NE system has been implicated in maze-learning behavior of rats by Anlezark, Crow, and Greenway (1973), who showed that bilateral LC lesions impaired learning and resulted in decreased cortical concentrations of NE. Neurophysiological evidence that NE modulates neuronal plasticity has been found recently in rat hippocampus by Hopkins and Johnston (1984) and earlier in cat visual cortex by Kasamatsu, Pettigrew, and Ary (1979).

In a series of experiments (see Kasamatsu, 1983, for a comprehensive review), Kasamatsu and his colleagues showed that when applied to visual cortex, 6-hydroxydopamine (6-OHDA), a neurotoxin that locally destroys LC projections, prevents the ocular dominance shift in MD kittens. They went on to show that plasticity could be restored with direct injections of norepinephrine. Kasamatsu et al. (1979) also attempted to induce plasticity in normally reared adult cats via NE injections. This experiment yielded an asymmetric U-shaped histogram biased toward the open eye. That is, NE was shown to produce observable plasticity in a system that otherwise would not demonstrate environmentally driven modification.

Attempts to replicate and extend these results have had mixed outcomes. Consequently, the role played by NE in modulating plasticity is very controversial at the present time. While NE certainly has *some* effect on the susceptibility of neurons to their stimulus environments, the situation is not as simple as one might think on the basis of Kasamatsu's initial reports. A study by Bear et al. (1983) exemplifies the mixed nature of the results and provides a partial explanation. They obtained catecholamine depletion using two different techniques to administer 6-OHDA. One group of kittens received 6-OHDA using the osmotic minipump procedure of Kasamatsu et al. (1979) and the other group received injections intraperitoneally (into the abdomen) before the third postnatal day. Plasticity was disrupted only in the former case; however, subsequent biochemical analysis (HPLC) revealed that the depletion was more effective in the latter group. The results were quite surprising: While the former group replicated the earlier result, the latter group showed no observable disruption of plasticity even though pharmacological analysis indicated that the NE depletion was more severe.[2]

Other examples of induced plasticity. Stimulation of at least three different sites has been shown to enhance plasticity in visual cortex of adult cats. Mechanical stimulation of the extraocular muscles generates

[2] By way of explanation, Bear et al. suggest that if the LC mechanism is disrupted sufficiently early, a compensatory mechanism may be activated. Such a mechanism, they point out, may have been discovered by Harik et al. (1981) in the rat.

a proprioceptive (feedback) signal which has been shown to be sufficient for plasticity with respect to both orientation selectivity (Buisseret, Gary-Bobo, & Imbert, 1978) and ocular dominance (Freeman & Bonds, 1979). Plasticity-gating capabilities as revealed by electrical stimulation techniques have also been reported for the reticular formation (Singer, 1979; Singer & Rauschecker, 1982) and the medial nuclei of the thalamus (Singer, 1982; Singer & Rauschecker, 1982).

The Effect of the Environment on the Critical Period

The conjecture that NE is a global modulator of neural plasticity does not necessarily imply that it governs the sensitive period. In this chapter an alternative conjecture is entertained; namely, that the potential for change possessed by a neuronal state is a function *of that state.* (It may perhaps be influenced by global factors as well.) The following result, mentioned earlier, of Cynader et al. (1976) provides supporting evidence. Kittens were raised in total darkness for periods as long as a year, after which rearing continued under the MD paradigm. Complete shifts to the open eye were found in each case, suggesting that visual input influences not only the response properties of the visual system by way of some plasticity mechanism, but the functioning of the mechanism itself. Ramachandran and Kupperman (1986) have taken these results further to show that the ocular dominance shift under MD is not the only plasticity phenomenon to be delayed (or prolonged) by dark rearing. They have not only obtained ocular dominance shifts in dark-reared animals, but they have observed dominance reversal upon reverse suture in animals at ages up to 6 and 6.5 months. The details of how much and what kind of experience are necessary and/or sufficient to delay the critical period are far from clear at this time. Mower, Christen, and Caplan (1983) report that 6 hours of binocular experience at 6 weeks eliminates the plasticity-preserving effect of dark rearing; but Ramachandran and Kupperman (1986) have shown that kittens raised normally for a full month and subsequently reared in darkness show plasticity in Area 17 for at least 6 months.

THE OCULARITY STATE AND ITS EFFECT ON PLASTICITY

The theoretical formulation in this chapter is intended to serve primarily as a partial explanation of the sensitive period for the development of ocular dominance in visual cortex. The central concept is that

retino-cortical connectivities are weaker in the newborn and are hence more susceptible to change than the stronger synaptic connectivities that develop as a result of experience. Assuming that the ocular dominance class depends on the relative synaptic weights (the *direction* of the synapse vector), this means that the effect on the ODH of a given change in connectivity is greater when the absolute values of the weights (the *magnitude* of the synapse vector) are relatively small. This notion can be supported by a very diverse variety of learning rules (i.e., synaptic modification rules), and so the specific formulae used here are secondary to the major theme. They are included for illustrative purposes.

The Model Neuron

Let the response r of an *ideal neuron* be given by a continuous nondecreasing function σ of its net depolarization x, and let x in turn be a weighted sum of the neuron's stimulus components s_i; i.e., the units have a *semilinear* (see Chapter 2) activation function:

$$r(t) = \sigma(x(t)) = \sigma\left[\sum_{i=1}^{N} w_i(t)s_i(t)\right]. \tag{1}$$

The weights w_i are collectively designated by the weight vector $\mathbf{w}$, and the stimulus elements s_i are collectively designated by the stimulus vector $\mathbf{s}$. Since the weights are associated with synaptic connectivities, $\mathbf{w}$ is called the *synaptic state*. One can clearly see that under this assumption the neuron's *response characteristics*, as measured by a set of test stimuli, are determined by $\mathbf{w}$. Adaptive learning (i.e., plasticity) is a dynamic process in which the synaptic state is driven by the stimulus environment such that the response characteristics with respect to the same environment approach some optimizing criterion. Several theoreticians (see Chapter 2) have analyzed the adaptive capabilities of various networks (idealized anatomical configurations) that self-organize according to rules (idealized physiological processes, usually at the scale of the synapse) that express the time rate of change of synapse weight w_i as a function v of local variables:

$$\frac{dw_i}{dt} = v(r(t), w_i(t), s_i(t), \cdots). \tag{2}$$

By hypothesizing an explicitly time-dependent factor into the above modification rule, the notion of a fixed critical period could be easily

incorporated. Global modulators such as NE can also be factored into the rule by appropriate independent functions.

However, as will be seen, no such ad hoc description is necessary. Rather, for certain initial conditions on $\mathbf{w}$, there exist modification rules from which plasticity decreases naturally under the influence of patterned input.

The Ocularity Plane

Figure 2 is a schematic view of the retino-geniculo-cortical pathway, the most direct (just two synapses) from the ganglion cells of the retina to the visual cortex. Note that the individual components of the stimulus vectors incident to a cortical cell are actually the activities of neurons in the lateral geniculate nucleus responding to retinal signals.[3] Despite findings that some LGN cells respond binocularly (Sanderson, Bishop, & Darien-Smith, 1971),[4] the components s_i are idealized as being strictly monocular. Let $C(I)$ be the set of all afferents to a given neuron from LGN neurons responsive to the contralateral (ipsilateral) eye. Let each neuron's *ocularity state* $\mathbf{W} \equiv (W_C, W_I)$ be given by its net input connectivities from each eye, and let the *ocularity plane* be the associated state space:[5]

$$W_C \equiv \sum_{i \in C} w_i \qquad W_I \equiv \sum_{i \in I} w_i. \tag{3}$$

[3] The retina is not the only neural structure that projects to the LGN. Other sources of LGN input include the reticular complex and feedback from Layer 6 of visual cortex (Gilbert & Kelly, 1975).

[4] Sanderson and his colleagues found that only about 9% of geniculate neurons in the cat could be excited from the nonpreferred eye, but that only 18% are truly monocular, the remaining majority having inhibitory fields in the nondominant eye. In monkey the LGN seems to maintain a higher degree of segregation; Marrocco and McClurkin (1979) report that 87% of neurons in macaque LGN are strictly monocularly driven.

[5] This is actually a special case of a more general formulation in which the ocularity state $\mathbf{W}$ is given by the net depolarizations of the neuron by arbitrary (fixed) monocular stimuli to each eye. Namely, this is the special case in which the test stimuli are uniform activations of all afferents. The typical laboratory procedure is quite different. In a given neuron the optimal stimulus is found for each eye. Then the two corresponding responses are compared in order to determine the ocular dominance class. Hence, the test stimuli are state dependent. A more detailed theory (such as Bienenstock, Cooper, & Munro, 1982) would be required to simulate this experimental method. However, the simple approach used here is adequate for the topic at hand.

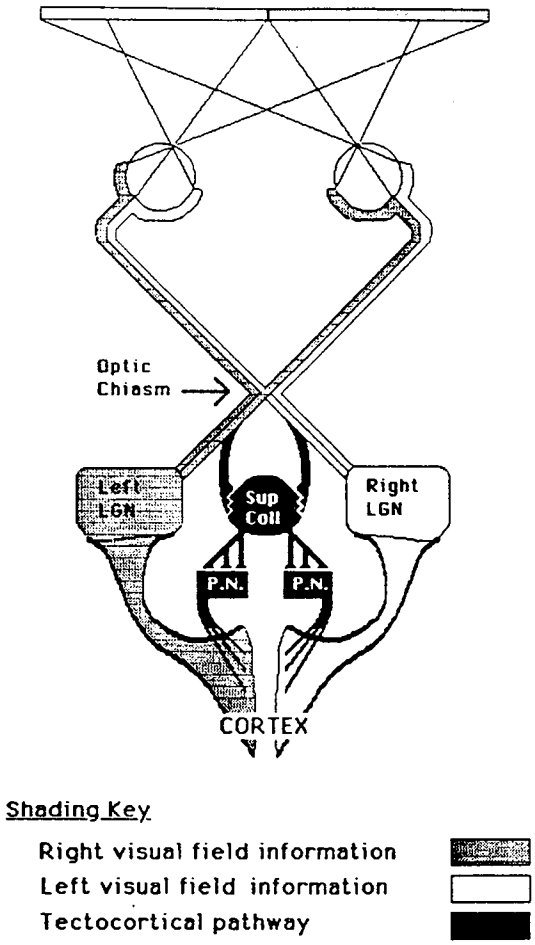

FIGURE 2. Circuitry of the visual system. Of the information pathways from retina to visual cortex, the retino-geniculo-cortical pathway, shown here in gray and white, is the most direct and the most well understood. The optic nerves approach the optic chiasm in two bundles, one associated with each eye. Here, they are separated and recombined into two groups according to whether they originate from the left or right side of the retina—i.e., they are grouped as a function of which "side" of the world they look at, such that information from the left side of the visual field is generally directed to the right side of the brain and vice versa. Other pathways from retina to cortex exist, for example, the tectocortical pathway: retina → superior colliculus → pulvinar nucleus → cortex (shown here in black).

Figure 3 depicts these assumptions with respect to a neuron in visual cortex. Macy, Ohzawa, and Freeman (1982) define a continuous measure of neuronal ocular preference, the ocular dominance index (ODI)

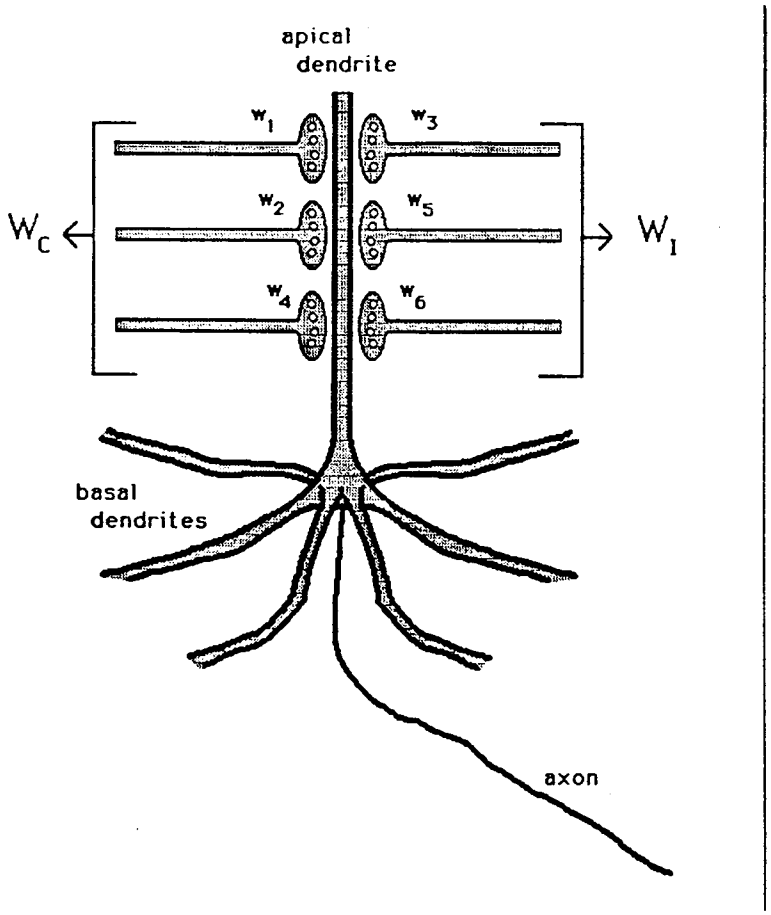

FIGURE 3. Schematic view of a neuron. Cortical neurons actually receive several types of inputs from several sources. The highlighted inputs represent monocular signals from LGN, though these do not necessarily project only to the apical dendrite as illustrated. These are segregated in the figure to illustrate the relationship between W and the synaptic weights w_i. In this example of just six synapses, $C=1,2,4$ and $I=3,5,6$. (In general these two sets need not have the same number of elements.) These synapses are thought to be purely excitatory. Other inputs, both excitatory and inhibitory, come from proximal cortical units, other areas of cortex, and other structures in the brain.

to be the quotient between the optimal ipsilateral response and the sums of the individual monocular optimal responses:

$$\text{ODI} \equiv \frac{r_{ipsi}}{r_{ipsi} + r_{contra}}. \tag{4}$$

Given *fixed* monocular test stimuli, the ODI can be roughly mapped onto the ipsilateral connectivity fraction z, a function of $\mathbf{W}$:

$$z(\mathbf{W}) \equiv \frac{\sigma(W_I)}{\sigma(W_I) + \sigma(W_C)}. \tag{5}$$

Using this index, the ocular dominance categories can be mapped onto regions of the ocularity plane (Figure 4). Two prominent aspects of the diagram are determined by the particular function σ. These are: (a) a threshold connectivity W_θ for both components of $\mathbf{W}$ below which no action potentials are elicited by the corresponding test stimuli and (b) the curvature of the iso-ocularity contours. (These are straight radial lines if σ is a linear function.)

The power of the ocularity plane as a tool in the theoretical analysis of ocular dominance plasticity will become clear presently. Scatterplots of experimental data on the ocularity plane may prove more valuable for presenting results than the ocular dominance histogram if accurate techniques for measuring $\mathbf{W}$ become available.

A hypothesis for the critical period. The introduction of the ocularity plane makes the task of describing the relationship between a neuron's state and its plasticity much easier than it would be otherwise.

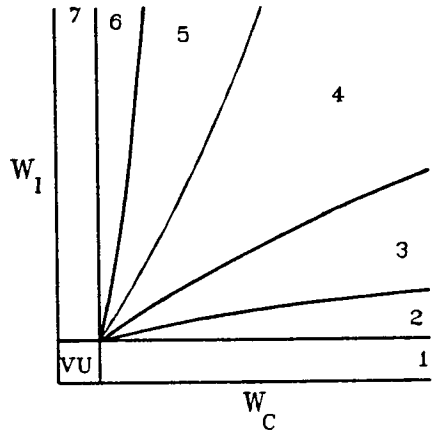

FIGURE 4. The ocularity plane. The plane is defined by axes corresponding to the net connectivities from both eyes to a cortical neuron. Loci of constant ODI are mapped onto contours that radiate from the point $\mathbf{W}_\theta \equiv (W_\theta, W_\theta)$, where the value W_θ corresponds to the firing threshold for some test stimulus. These have been chosen here to mark arbitrary boundaries that might be used to delimit the seven ocular dominance classes defined by Wiesel and Hubel (1963). Running orthogonal to these contours is, by definition, the gradient field for ocular dominance.

This description is repeated in the next section using more formal mathematical language for those who are interested, but the gist can be understood quite well without wading through the mathematics. The arguments in this chapter are founded on the following assumptions:

A1. Initially the synaptic weights are small.

A2. As it is molded under the influence of a structured stimulus environment, the pattern of synaptic weights is refined and reinforced with experience, and thus tends to grow.

A3. The rate of change of synaptic weights may decrease or remain invariant, with changes in the magnitude of the weight vector; in fact, the rate of change of synaptic weight may even *grow* as the magnitude of the weight vector grows, as long as it grows less rapidly than the weight vector itself.

These assumptions are readily understood in the context of the ocularity plane. Since the initial weights are small (A1), the ocularity state is initially near the point where the ocular dominance classes converge ($\mathbf{W}_\theta$). In this region, plasticity is high because it requires only a small change in the synaptic weights to move the state from one OD class to another. As its weights increase (A2), the neuron's ocularity state is driven away from the origin where increasingly larger weight changes are required to produce shifts in ocular dominance class. Assuming the ability to change weights does not also grow in proportion to the growth of the weight vector (A3), this results in an observed loss of plasticity.

State Dependence of Plasticity

Given z, a scalar measure of the neuronal state, the plasticity p associated with a particular state W can be defined for any stimulus environment E. The plasticity is the time rate of change in z under the influence of E:

$$p \equiv \left\langle \frac{dz}{dt} \right\rangle_E \equiv \frac{\partial z}{\partial W_C} \bar{V}_C + \frac{\partial z}{\partial W_I} \bar{V}_I \equiv \nabla z \cdot \bar{\mathbf{V}} \qquad (6)$$

where $\bar{\mathbf{V}} \equiv \langle d\mathbf{W}/dt \rangle_E$ is the average "velocity" of the state in the W-plane and ∇z is a vector pointing in the direction of steepest change in z with magnitude equal to that change (the components of ∇z are simply the partial derivatives of z with respect to W_C and W_I). The plasticity can thus be represented as the inner (dot) product between two vectors: the velocity of the state and the gradient of the ODI. The

inner product of two vectors can be expressed as the product of their magnitudes and the angle between them (see Chapter 9). These are considered separately below.

Dependence on the magnitudes of ∇z *and* $\bar{\mathbf{V}}$. Whether the loss of plasticity associated with the critical period is a function of neuronal state ultimately depends on the behavior of the product between the magnitudes of these two vectors for large magnitudes of $\mathbf{W}$. Since $|\nabla z|$ decreases with $|\mathbf{W}|$ ($|\nabla z|$ goes as $|\mathbf{W}|^{-1}$ on average), the issue reduces to the relationship between $|\bar{\mathbf{V}}|$ and $|\mathbf{W}|$. This is exactly the issue addressed in Assumption A3 above, since $|\mathbf{W}|$ is the magnitude of the weight vector and $|\bar{\mathbf{V}}|$ is its speed in the ocularity plane. In keeping with the conventional Hebb-like approach (i.e., the change in each synaptic weight depends only on the activities of the corresponding presynaptic and postsynaptic units), consider the indirect effect of $\mathbf{W}$ on $\mathbf{V}$ mediated by the neuronal activity rather than any explicit dependence of $\mathbf{V}$ on the weights. It now becomes very important to consider the form of the function σ relating the response r to the quantity x (see Equation 1). This function is sometimes called a *squashing* function since it tends to saturate for high values of x (as shown, for example, by Chapman, 1966) because of the upper limit on the neuronal firing rate (i.e., the absolute refractory period). Hence, if $\bar{\mathbf{V}}$ depends on r (not x), it will (like r) level off as x increases.

It would be sensible for $\bar{\mathbf{V}}$ to *decrease* with increasing x. The argument supporting this possibility is based *teleologically* on the assumption that *each unit optimizes some function of its stimulus environment according to a gradient descent process*. In Chapter 8, it is shown that optimization of least square error in a multilayered system leads to an explicit rule for learning in which each synapse modifies according to a function separable into three factors: the presynaptic (afferent) activity, an error signal propagated from higher levels, and *the derivative of the squashing function of the postsynaptic unit*. More generally, one can show that the derivative $\sigma'(x)$ is a factor regardless of the function being optimized, whether it be a least square error (as in Chapter 8); selectivity for orientation, direction, or spatial frequency (as in many cells in visual cortex); or some other quantity. This is demonstrated quite easily. Let Q be the optimized function, a quantity measured over the set of response values r^i evoked from a given unit by a set of test stimuli $\mathbf{s}^i$. If $r^i \equiv \sigma(x^i)$ (where $x^i \equiv \mathbf{w} \cdot \mathbf{s}^i$), then the gradient ∇Q breaks down as follows:

$$\frac{\partial Q}{\partial w_j} = \sum_i \frac{\partial Q}{\partial r^i} \frac{dr^i}{dx^i} \frac{\partial x^i}{\partial w_j} = \sum_i \frac{\partial Q}{\partial r^i} \sigma'(x^i) s_j^i. \tag{7}$$

With increasing exposure to the world, units may become highly selective so that they come to respond at intermediate levels where $\sigma'(x)$ is high (see Figure 5) less frequently, and may hence become less plastic with respect to that environment.

The first derivatives of certain squashing functions can be expressed as functions of the squashing functions themselves—i.e., they satisfy $\sigma'(x) = F(\sigma(x))$. A simple (i.e., easily implemented) function F gives a technical advantage that may be exploited at two levels: (a) by the biochemical mechanisms that underlie plasticity and (b) by the

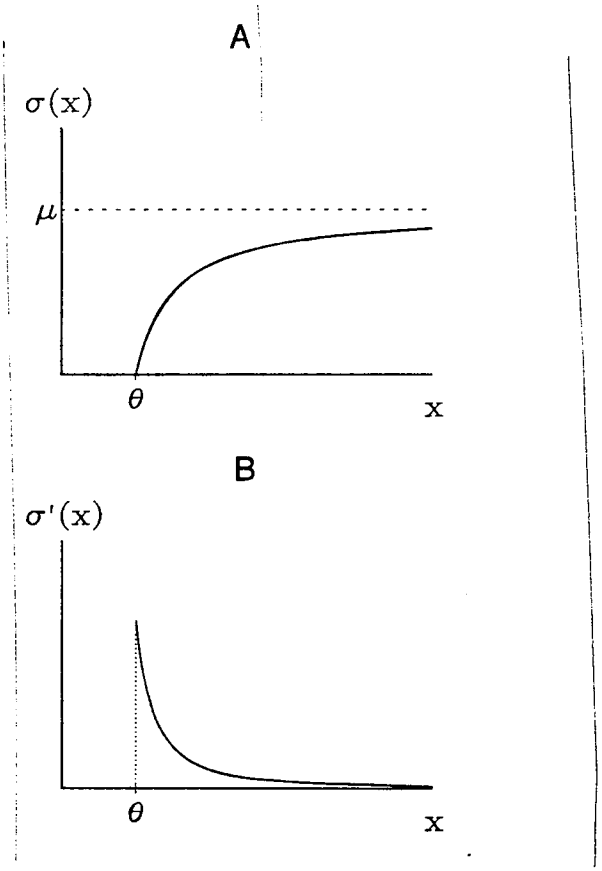

FIGURE 5. A squashing function and its first derivative. *A*: An easily computed function that incorporates the salient features (namely, a threshold, θ, and monotonic approach to an asymptote, μ) of neuronal response characteristics is given by the function that is defined by $\sigma(x) \equiv \mu(x-\theta)/(x_{1/2}+x-2\theta)$ for $x > \theta$ and vanishes for $x < \theta$, where $x_{1/2}$ is the value of x at which the function attains half its maximum value (i.e., $\sigma(x_{1/2}) = \mu/2$). *B*: The derivative $\sigma'(x)$ of this particular function is shown for $x > \theta$. The decay in the derivative as x increases can be implicated in a decay in learning rate for systems that follow gradient descent. This effect is potentially quite substantial as indicated by the very small values of the derivative.

computer algorithms used to simulate them. A few such pairs of functions and derivatives are given in Table 2 along with the function F that expresses the function relating each pair.

Dependence on the angle between ∇z *and* $\overline{\mathbf{V}}$. Analysis of the time course of the OD shift under MD (Blasdel & Pettigrew, 1978) has shown that the binocular OD classes are depleted before those associated with the closed eye, resulting in a U-shaped OD histogram. Since the inner product between the vector fields (and hence the observed plasticity p) tends to increase as the fields are more *parallel*, the elevated plasticity in the binocular regions suggests that the fields may be more parallel in these regions than in the monocular regions of the ocularity plane. Since the shape of the ∇z curves is known to be somewhat circular about the origin, or *concave inward*,[6] the approximate shape of the $\overline{\mathbf{V}}$-field can be deduced. It follows the $\overline{\mathbf{V}}$-field trajectories are directed radially near the axes and oriented tangentially in the binocular regions; i.e., they are *concave outward* (Figure 6A). The resulting elevated plasticity level for the binocular states can be seen in Figure 6B which illustrates the plasticity function in the context of a stimulus environment restricted by monocular deprivation (i.e., $E = MD$) using a simplified version of the Bienenstock et al. (1982) formula for $\overline{\mathbf{V}}$. Note that the states furthest from equilibrium (those at the "beginnings" of the trajectories) are initially driven *toward* the origin, but are "deflected." A useful analogy (suggested by a remark of Paul Smolensky's) is that the origin "repels" states under the influence of a patterned environment. It is thus seen that while the states *tend* to grow, sudden changes in the statistics of the visual environment (e.g.,

TABLE 2

SAMPLE SQUASHING FUNCTIONS AND FIRST DERIVATIVES

$\sigma(x)$	$\sigma'(x)$	σ' in terms of σ
$\frac{x}{1+x}$	$\frac{1}{(1+x)^2}$	$\sigma'(x) = (1-\sigma)^2$
$\frac{1}{1+e^{-x}}$	$\frac{-e^{-x}}{(1+e^{-x})^2}$	$\sigma'(x) = \sigma(1-\sigma)$
$1-e^{-x}$	e^{-x}	$\sigma'(x) = 1-\sigma$

[6] Concaveness can be assigned a direction corresponding to the second derivative (acceleration) of a trajectory.

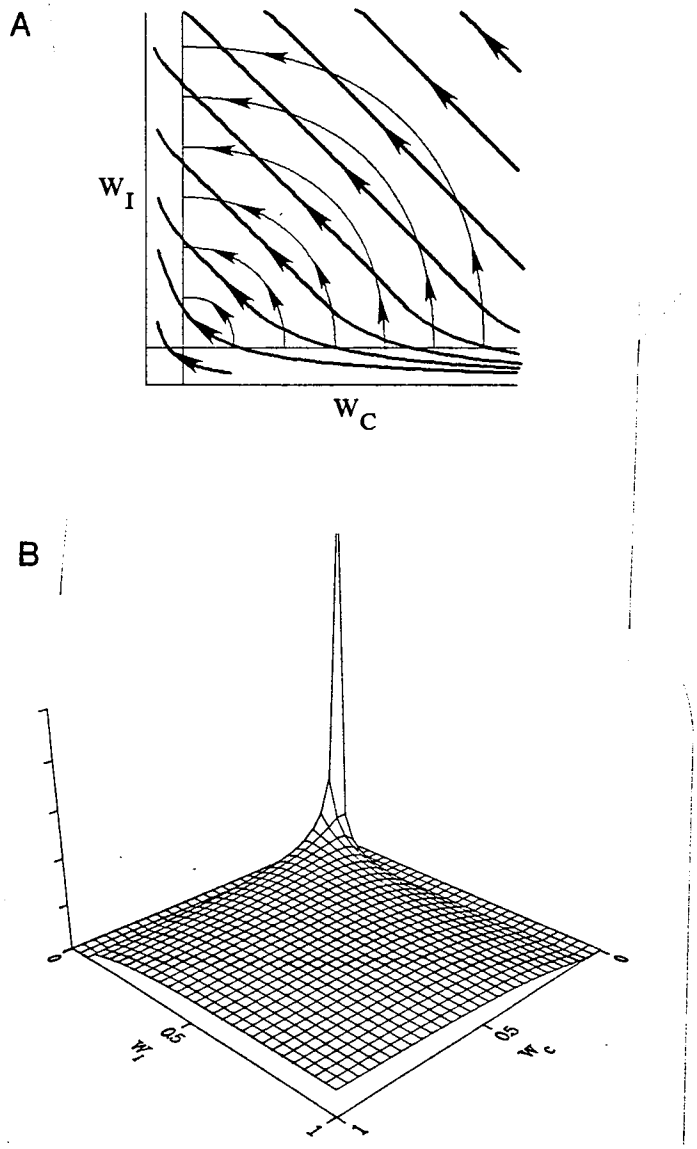

FIGURE 6. Plasticity in a monocularly deprived environment as a function of the ocularity state. *A*: The velocity field **V** (bold) is shown together with the gradient field ∇z. *B*: Plasticity is plotted as a function of the state **W**. Note that the origin is located in the *far* corner of the base plane so that the region obstructed by the large peak is minimal.

RS) may deliver an "impulse" in the direction of the origin which is eventually overcome by the system dynamics to drive the state outward, most likely in a direction different from its approach.

These last two paragraphs have formally expressed a notion that is actually simpler than the mathematical nature of the presentation might indicate. The fundamental idea is that the *observed* plasticity depends on the objective change in the neuronal state (i.e., $\mathbf{V}$) *relative to the plasticity measure* (in this case, the ocular dominance index z). Aspects of this relative change that have been considered are the absolute magnitude $|\overline{\mathbf{V}}|$ of the change in neuronal state, the sensitivity $|\nabla z|$ of the plasticity measure, and the angle between the change in the state and the contours of constant z.

COMPARISON WITH EXPERIMENTAL DATA

Developmental models of visual cortex have for the most part focused on the final (equilibrium) states inherent in the dynamics of hypothetical systems (e.g., Amari & Takeuchi, 1978; Bienenstock et al., 1982; Cooper, Lieberman, & Oja, 1979; von der Malsburg, 1973), ignoring the variability in the convergence rates of these systems that results from abrupt changes in the visual environment such as those manipulations used to test plasticity. Ocularity plane analysis is an attractive technique in the design of a detailed rate-descriptive model for learning.

The time course of the reverse suture paradigm. Qualitative analysis of the *shape* of the trajectories (as in Figure 6A), however, leads to a prediction for the RS paradigm. It appears that early reversal of monocular deprivation should bring the states closer to the origin as they migrate from the neighborhood of one axis toward the other axis (see Figure 7). Hence they might easily go from one monocular class to the other via the VU region—i.e., pass between the origin and the plasticity singularity $\mathbf{W}_\theta$. Consider states which begin further from the origin on the axis associated with the newly closed eye. After reversal, their paths are more likely to traverse the binocular regions of the ocularity plane. Therefore the intermediate (U-shape) stage of RS is expected to be less sharp, i.e., there should be more binocular units observed if the reversal is performed relatively late (but before plasticity vanishes).

The time course of the ocular dominance shift under RS has been analyzed by Blakemore and Van Sluyters (1974). As in the early stages of MD rearing, the OD histogram initially becomes somewhat U-shaped. Casual consideration of this in terms of the $\mathbf{W}$-plane leads to the conclusion that the trajectories must be strongly concave outward, matching the result derived in the previous section for simple MD. One explanation for these results rests on the assumption that many

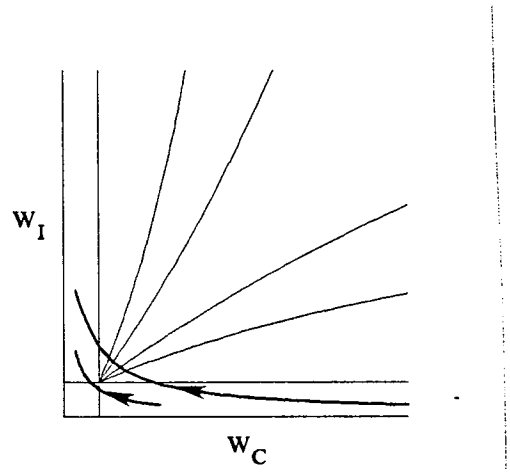

FIGURE 7. Trajectories corresponding to reverse suture. Two examples are shown. One path (that closer to the origin) starts from a weaker, and hence presumably less mature state. This state travels from Class 1 to 7 without entering the intermediate classes.

units maintain subthreshold synaptic input from the closed eye during monocular deprivation for some period of time.[7] The possibility that these inputs eventually disappear presents another candidate for the loss of plasticity by monocular deprivation. Computer simulation results of the Blakemore and Van Sluyters (1974) experiment are shown together with the experimental data in Figure 8.

Alternating monocular occlusion. Consider a stimulus distribution E corresponding to the experimental paradigm for alternating monocular occlusion. If the alternation is sufficiently frequent, states in the neighborhood of the diagonal line $W_C = W_I$ should display low plasticity since they are driven along that line, if at all; hence they move in a direction orthogonal to the ODI gradient (i.e., along a contour of constant ODI). This is illustrated by the plasticity function in Figure 9 and accounts for the W-shaped OD histograms (i.e., Columns 1, 4, and 7) obtained by Blasdel and Pettigrew (1979) in some of their alternating deprivation studies.

How is connectivity related to responsivity? A scatter-plot of a sample of cortical neurons in the ocularity plane adds a dimension to the

[7] Evidence supporting the existence of such input has been found in cat visual cortex by sudden removal of the open eye after a period of monocular deprivation (Blakemore, Hawken, & Mark, 1982; Kratz, Spear, & Smith, 1976), but Dräger (1978) reports the opposite finding for mice.

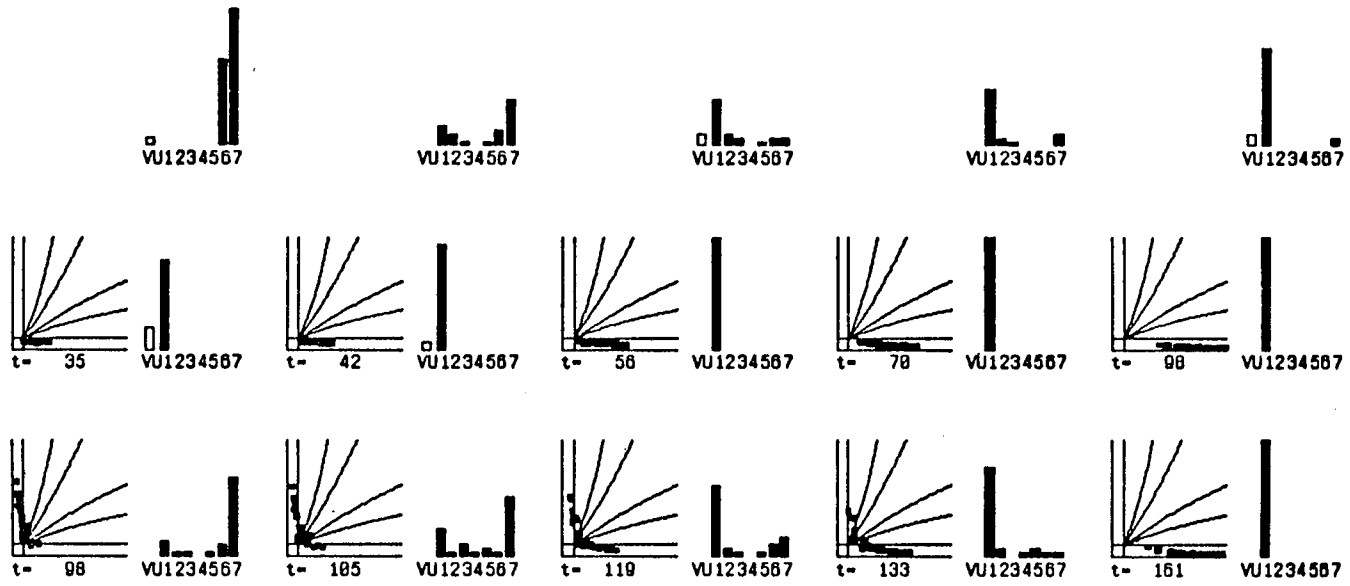

FIGURE 8. Simulation of reverse suture experiments run by Blakemore and Van Sluyters (1974). In each case, animals were deprived in one eye for some variable period (35 days, 42 days, 56 days, 70 days, 98 days) and then subjected to a suture reversal which was left in place for 63 days before recording. *Top row:* Experimental results. *Middle row:* Simulation state at time of suture reversal. Scatter plots in the ocularity plane are shown with the corresponding ocular dominance histograms. *Bottom row:* Simulation state at time of recording (compare with top row).

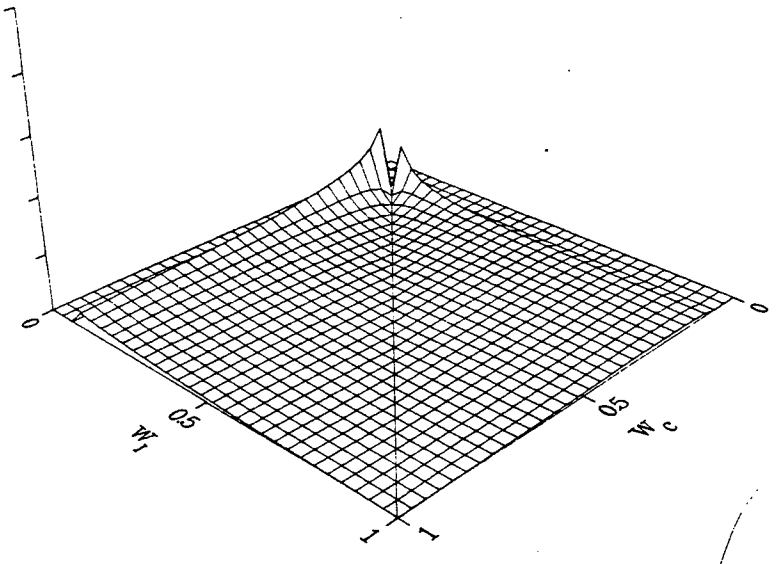

FIGURE 9. Plasticity under alternating monocular occlusion. Note the lack of plasticity along the diagonal. In an AM environment, highly binocular units are not as plastic as units with a bias for one eye. They are effectively "torn" between committing themselves to one eye over the other.

information in the ocular dominance histogram. In their study, Freeman, Sclar, and Ohzawa (1982) produce a similar representation which maps units onto a ODI vs. responsivity plot. These can be thought of as the basis of another (polar-like) coordinate system in the ocularity plane (see Figure 10). The ocularity plane would be much more viable as an analytic tool if a precise relationship could be established between synaptic connectivity and responsivity. A finer measure of response strength might be accomplished using intracellular electrodes to count individual EPSPs (and perhaps IPSPs as well). Lacking this, an objective and more sensitive measure of the degree of response should be adequate to the extent that it ought to correlate well with plasticity. Not only should the *degree* of response be included in such a measure, but such an index might incorporate several aspects of the response including not only the firing frequency, but also (for example) the latency of firing.

The role of norepinephrine: A conjecture. The prolongation of the critical period induced by BD observed by Cynader and the norepinephrine effect reported by Kasamatsu and Pettigrew (1979) extend

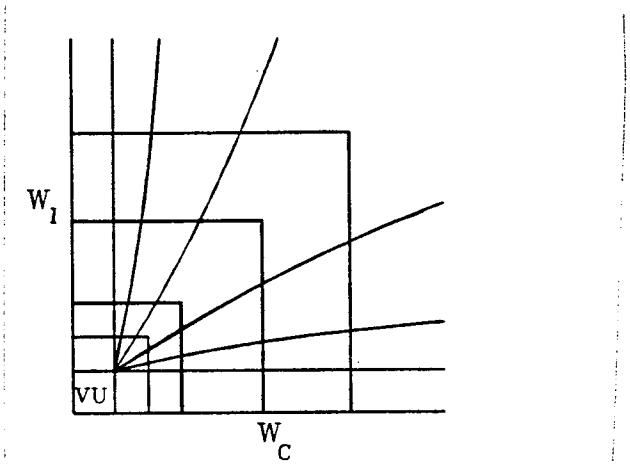

FIGURE 10. Ocular dominance and responsivity as polar-like coordinates. Accurate measures of W may be difficult to come by. One can only calculate rough approximations on the basis of the corresponding responses. Some estimates might be made on the basis of existing data by using ocular dominance as an indication of the "angle" and responsivity as a rough measure of the "magnitude" of W.

the concept of critical period far beyond the original work by Hubel and Wiesel. Of the many differences between these two approaches, there is one which is fundamental to the nature of learning by biological systems—namely, that while the rate of learning in biological information-processing networks can be made subject to *extrinsic* factors, it may well be influenced by *intrinsic* properties of the learning mechanism, the information patterns as they are learned, and interactions between them.

It has been established in this chapter that a theory of plasticity need not presuppose the existence of a global modulator to account for the critical period. What role, then, does the norepinephrine system play? After all, it has been well demonstrated, both experimentally and on teleological grounds, that it can modulate plasticity.

The conjecture that the LC releases NE as part of an attentional mechanism[8] seems to resolve the issue. Rather than acting as a clock controlling the developmental sequence of plasticity, the activity of NE probably fluctuates on a finer time scale. Under this assumption, nonrecognition of a stimulus might trigger a "novelty alarm" in the form of an LC/NE plasticity signal. Thus, another factor in the critical period

[8] The notion that NE is a medium for externally gated plasticity originated in the early 1960s based on studies of the reward system (Poschel & Ninteman, 1963). Whether and how catecholamines are implicated in reward and reinforcement remains controversial. (For a review, see Wise, 1978.)

may be the loss of novelty brought about by experience; this suggests an explicit experience-dependent mechanism for plasticity—one that would *amplify* the effect described in this chapter and would therefore be difficult to analyze independently.

Grossberg (1980) offers a comprehensive description of the critical period that incorporates all of the mechanisms in Table 1. This is quite different from the approach taken in this chapter, in which I have attempted to separate the critical period from these other factors. Given the broad disparity between these two points of view, an experiment to discriminate between them should not be difficult to design. For example, what are the states of these mechanisms in cats reared in darkness for several months from birth?

DISCUSSION

State Dependence and Existing Modification Rules

Assumptions A1, A2, and A3 can be incorporated into many dynamical models of plasticity in cortex, as can hypotheses based on global modulation (see next subsection). Consider three varieties of mathematical formulations, characterized by the factors determining the magnitudes of the vectors representing the final equilibrium states (see Figure 11):

- *Case 1: Throughout modification, the magnitudes of the neural states are explicitly constrained; for example, the sum of the components is held constant* (e.g., Perez, Glass, & Shlaer, 1975; von der Malsburg, 1973; Chapter 5). Unfortunately, explicit constraints on magnitude are not consistent with the assumption that the state vectors grow during learning. Certain models of this kind employ such constraints because of a tendency for the weights to grow without limit; for these models, only an *upper bound* is really needed and the inherent tendency for growth may provide the necessary reinforcement. However, some models do require that strict magnitude constraints be imposed (e.g., those with even less stable dynamical equations). While it may not be impossible to adapt these models to Assumptions A1, A2, and A3, the solution is not likely an elegant one; for example, a requirement that the magnitude be conserved at a fixed value M might be altered such that M increases according to some complex ad hoc rule up to some maximum value M_{max}.

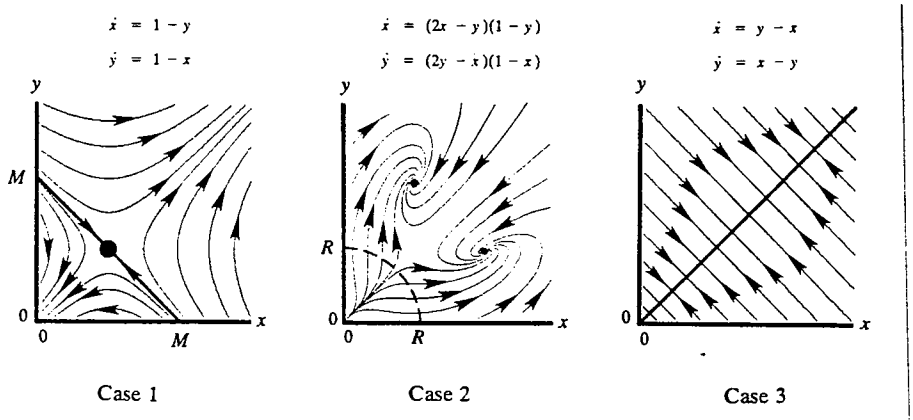

FIGURE 11. Three kinds of dynamical systems. These three examples (instances of Cases 1, 2, 3 in the text) illustrate certain important characteristics of models relevant to the assumptions stated earlier for state-dependent plasticity. The dynamical systems used are given above each graph in terms of two generic variables, x and y. *Case 1*: In this case, the trajectories are driven to either the origin or to infinity, unless the constraint $x+y = M$ is imposed. *Case 2*: Here, there are two global equilibrium points. Thus, under the assumption that the initial states are close to the origin (e.g., within the dotted circle $x^2+y^2 < R^2$), the states will tend to grow. *Case 3*: All states are driven to the nearest point satisfying $x = y$ and hence all trajectories tend to preserve the sum $x+y$ of the initial state. Thus, there is no tendency for the states to grow.

- *Case 2: While no explicit constraints are placed on the weights, the system dynamics drive all states outside some nonzero radius R about the origin* (e.g., Amari & Takeuchi, 1978; Bienenstock et al., 1982). Models of this class are most readily adapted to Assumptions A1–A3 since they have a "natural" tendency to drive states with sufficiently small magnitudes away from the origin. The value R (like M in Case 1) gives a scale for selection of the initial weights—i.e., "small" in A1 is interpreted here as small relative to R. Three kinds of behavior are possible in the limit $t \to \infty$; these are (a) convergence to an equilibrium state with magnitude greater than R, (b) oscillations (e.g., limit cycles) that come no closer to the origin than R, or (c) divergence. It must be emphasized that such models must be carefully tuned if they are to give an accurate account of learning rate. This tuning is not necessarily limited to selecting parameters for the learning rules; good matching to the data may require alterations of the form of the rules themselves such as the derivative factor in Equation 7.

- *Case 3: There is no minimum radius such as that in the previous case; equilibrium states exist arbitrarily close to the origin* (e.g., Munro, 1984). A dynamical system of this variety presents a particularly hostile context for Assumptions A1–A3. There is no reason to expect the weights to grow no matter how small the initial weights are. This highlights the fundamental dependence of the assumptions on *scale*.

Global Modulators as Dynamical Variables

Global modulation of cortical plasticity, whether it is mediated by the locus coeruleus, proprioception, the brain stem, or the thalamus, can be integrated into a synaptic modification formula in any of several ways. Let us assume a general modification rule that is separable into a presynaptically dependent function ψ and a postsynaptically dependent function ϕ:

$$\frac{dw_i}{dt} = \psi(s_i)\,\phi(r). \tag{8}$$

Some simple conjectures as to the role of a modulatory signal (α) are that it may influence plasticity in a direct multiplicative fashion or by exerting an effect on the presynaptic or postsynaptic activities:

$$\frac{dw_i}{dt} = \begin{cases} \alpha\,\psi(s_i)\,\phi(r) \\ \psi(s_i(\alpha))\,\phi(r). \\ \psi(s_i)\,\phi(r(\alpha)) \end{cases} \tag{9}$$

The first of these possibilities is consistent with the data from both Cynader's and Kasamatsu's labs. That is, a global modulatory signal may control the speed (magnitude of the velocity vector) of the neuronal state as it is driven by the environment, but the distance between ocular dominance classes depends on the state. Thus the observed plasticity is a function of *both* the synaptic state and global factors—e.g., norepinephrine. Destruction of the catecholamine terminals in some region by 6-OHDA might then bring synaptic modification to a complete halt.

Secondary global modulatory action via influence of the presynaptic activity is not a very promising approach since it would distort the pattern of afferent activity. This effect would be disastrous except in the trivial case, namely, if it were equivalent to a uniform scale factor. Of course, the action would then be formally identical to the third

conjecture for the role of α, in which it influences the postsynaptic activity (assuming that r depends monotonically on a variable that is linear in s_i, like x in Equation 1).

Norepinephrine is known to influence responsivity in neurons of the visual cortex. Kasamatsu and Heggelund (1982) found that this influence was not easily characterized; for example, overall responsiveness would increase, decrease, or not change with roughly equal probability. However, they did notice certain correlations with respect to cell type. Videen, Daw, and Rader (1984) also found that the relationship between NE and neuronal response is quite complicated. Thus the third of the above functions may have some support, though the exact form of the dependence of α on NE concentration has not been specified.

Predictions

Theoretical neuroscientists are often criticized (often with good reason) for neglecting their duty to suggest experiments that test their theories. They counter with accusations, which also are valid, to the effect that the bulk of theoretical work passes through the literature unnoticed by the experimental community and that many, if not most, predictions remain untested. Arguments assigning blame aside, the interaction of theory and experiment in neuroscience is not in a healthy state. In an effort to improve the situation, this section outlines some crude designs that may lead to laboratory verification of the ideas covered in this chapter.

A very informative experiment would be to analyze the time course of the loss of plasticity along the path of information flow (Layer 4c vs. other layers in the monkey) after a period of dark rearing. By comparing this result with the observation that Layer 4c loses plasticity for ocular dominance earlier than other layers in monkey Area 17 (Blakemore, Garey, & Vital-Durand, 1978; LeVay, Wiesel, & Hubel, 1980; Vital-Durand, Garey, & Blakemore, 1978), it should be possible to separate the maturational from the environmental influences on the critical period. A strict interpretation of the analysis in this chapter predicts that the entire process should be postponed and hence the results after dark rearing should not substantially differ from the established results. The outcome of the experiment could conceivably be mixed indicating that both environmental/informational and maturational factors exert influence. It seems reasonable, indeed quite likely, that at lower levels (such as the thalamus), maturational effects on the critical period outweigh the effect described in this chapter.

Reinstating plasticity after patterned rearing. Cynader's experiments have demonstrated that the visual system is capable of "waiting" in a plastic state for patterned input. This result may be indicative of a more general relationship between responsivity and plasticity.[9] One idea that grows out of the present approach is that plasticity may in fact be reduced in part because cells in the mature animal rarely respond in the intermediate range, where the derivative of the cell's response with respect to its net input is relatively large (see Figure 5B). If, as suggested earlier, synaptic modification depends on this derivative, then cells that are responding vigorously will not modify their connections. Thus these ideas predict that if the cells could be made to respond less vigorously, they could thereby become more plastic, provided the new response level were in the appropriate range. Of course, this is not the only factor that determines plasticity; the size of the weight vector plays a very important role in the extent to which a change of a fixed size will alter the ODI. That is, of the two quantities ∇z and $\sigma'(x)$ that are candidates for the loss of plasticity, only the latter can be manipulated in the adult. Hence, the observed plasticity resulting from a reduction of response magnitude in adult animals should be weaker than the natural plasticity of younger animals.

The following outline sketches an experimental paradigm for testing the relationship between responsivity and plasticity: One method to reduce responsivity without introducing the potential pitfalls of pharmacological techniques would be to apply diffusive contact lenses. Degrading the contrast of the stimuli has the further property of reducing orientation selectivity and hence simulates this aspect of kittenhood as well. This has the advantage of further simulating properties of the young animal and the disadvantage of confounding orientation selectivity and neuronal responsivity as factors in neural plasticity. Once the lenses have been calibrated with respect to neuronal responsivity, many different rearing paradigms should be tried since various factors may induce differential effects across different paradigms. Differences in morphological development (at both geniculate and cortical levels) are, for example, typically observed for different kinds of rearing; it may therefore prove advantageous to study the effects of manipulating the environments of normally reared animals since they have not suffered

[9] Cynader plays down the generality of this relationship, citing his observation that neuronal responsivity reaches adult levels before the peak of the critical period for animals raised in patterned environments. However, it should be noted that synaptic weights can continue to grow even after adult levels of responsivity are observed. This will naturally be the case in units whose activation saturates according to the general types of activation functions considered here and throughout this book.

the potentially irreversible and possibly plasticity-limiting damage that is associated with deprived rearing.

Regardless of the experimental technique, a clear demonstration that links weak neuronal responsivity to increased plasticity promises to both provide evidence that neuronal response characteristics evolve such that they follow a gradient descent with respect to some environment-related variable and lend strong support to the notion that plasticity is state dependent.

BEYOND VISUAL CORTEX

The ideas which have been described with respect to ocular-dominance plasticity in visual cortex are abstracted in this section and considered as they might apply to higher (more cognitive) levels of cortex. In general, it seems plausible to assume that the important informational content of a stimulus pattern in cortex depends upon the *relative* values of its components, which define radially oriented regions of the state space, since nearly every level of neural processing shows evidence of lateral inhibition, a known mechanism for pattern normalization. The absolute magnitude of the stimulus loses its relevance in this case and its informational content is reduced by one degree of freedom to the *direction* alone. Gati and Tversky (1982) refer to the magnitude and direction of a vector stimulus as its *qualitative* and *quantitative* aspects, respectively. One can clearly see that in this case the assumptions A1, A2, and A3 may apply under these circumstances. According to this distinction, the plasticity of a self-organizing process is expected to exhibit state dependence as a function of the *quantitative* aspect of the state across a *qualitative* measure on the stimulus environment. Small assemblies or modules of such units can be made to evolve to arbitrary regions of the input space (as in Reilly, Cooper, & Elbaum, 1982), which can be thought of as categories.

Stages of Processing

Environmentally driven sequential development. The state dependence of plasticity has interesting implications for a hierarchically organized system. Consider an organizational scheme in which information passes through a sequence of several neural assemblies, each

responsible for some stage of processing.[10] Assume that synaptic connectivities are initially weak throughout the entire system. Each stage lies in a state of weak connectivity and high plasticity until it receives sufficient exposure to some pattern set. At the onset of exposure to patterned input (e.g., eye-opening in kittens), only the input assemblies receive strong signals. The earliest stages in the system are therefore the first to lose plasticity, and so increasingly longer sensitive periods may be expected for the development of higher cognitive functions.

This consequence of the ocularity plane framework is consistent with data from visual cortex of both cat and monkey. Daw and Wyatt (1976) found evidence that the sensitive period for direction-selective cells is earlier, both in its onset and offset, than the sensitive period for ocular dominance. While these may not represent different stages of processing, the experiment clearly supports the notion of a link between plasticity and the nature of the stimulus features.

In monkey (and in cat as well), the LGN projection to Area 17 terminates most predominantly in Layer 4c (Hubel & Wiesel, 1972); this layer therefore receives the earliest, most "raw" input to cortex. Considerable evidence, both physiological (Blakemore et al., 1978) and morphological (Vital-Durand et al., 1978) indicates that in monkey visual cortex, plasticity is diminished in this layer earlier than in the other layers. These findings have been replicated by LeVay et al. (1980), who attribute the effect to maturational factors rather than implicit aspects of plasticity.

Based on developmental data from the rat somatosensory system, Durham and T. A. Woolsey (1984) have formulated a "sequential instruction hypothesis for sensory pathway pattern development" (p. 445). In spite of their conjecture that "a given station might 'instruct' the next more central one" (p. 425), they stick to a more traditional interpretation of their data: ". . . critical periods are linked to developmental periods since they coincide with times during which axons are normally innervating their targets and are segregating" (p. 446). It is not clear therefore whether the instruction to which they refer is exogenous (i.e., sensory information) or endogenous (e.g., genetically precoded chemical transmission).

Another factor in considering the developmental aspects of different levels of the system is the structure of the pattern environments presented to each stage. As successive levels self-organize, the representation of the world becomes more symbolic. During its ascent

---

[10] The visual system seems to be organized in such a fashion. However, the information flow is not strictly serial, rather it seems to be directed along "functional streams" (Van Essen & Maunsell, 1983), which have mixed hierarchical/parallel attributes.

through the system, information is continually recoded in a manner such that the stimulus sets become more discrete. Shepard and Podgorny (1975) make this point with the following example:

> Whereas we can continuously shift a color (for example, blue) in brightness, hue, and saturation until it becomes as similar as we wish to any other color (for example, green), we cannot continuously deform a word 'blue' to another word 'green' without passing through intermediate configurations that are not words at all. (pp. 189-190)

This observation pertains to plasticity in that as stimulus sets become more discrete, and perhaps more orthogonal, their constituents are better separated; hence learning by feature-detecting units is much more rapid. So it is conceivable that for such systems, the equilibrium points may not be so far from the origin; thus, these neurons may enjoy extended or indefinite sensitive periods! This, of course is consistent with our experience; people learn to recognize patterns quite rapidly (relative to the hours or days required for neurons to become orientation selective, for example) and do not generally lose that ability.

SUMMARY

In this chapter I have proposed a factor that may account for the high rate of learning observed in the early stages of many learning processes. This factor of state dependence may in fact be the principal component responsible for the critical period; it emerges from three basic assumptions regarding the mathematical relationships between synaptic connectivity values and their time derivatives. The notion of state-dependent learning rate has been discussed here in the context of the observed susceptibility of ocular dominance characteristics to environmental manipulations. This approach suggests an account for several specific observations in visual cortex and also suggests further experimental tests of the relationship between ocularity state and plasticity.

On a broader level, it is hoped that this chapter will contribute to greater interaction between theory and experiment in this field. Toward this end, analytical tools such as the ocularity plane may prove useful both for making experimental data more accessible to theoretical analysis and for examining and describing aspects of the mathematical models.

ACKNOWLEDGMENTS

Over the past three years, several individuals were very generous with their time and energy to apply their knowledge and insight to the issues raised in this chapter. These include Leon Cooper, Michael Paradiso, and Alan Saul of Brown University; Paul Heggelund of the University of Oslo; V. S. Ramachandran of UCSD; and several members of the PDP research group at the UCSD Institute for Cognitive Science, in particular, David Rumelhart, Paul Smolensky, and especially Jay McClelland for his thoughtful and meticulous proofreading. This work has been supported by United States Office of Naval Research Contract N00014-81-K-0136, the Norwegian Council for Science and the Humanities (NAVF), the System Development Foundation, and grant PHS MH-14268-11 from the National Institute of Mental Health to the Center for Human Information Processing at the University of California, San Diego.