Neural Networks, Vol. 5, pp. 403–413, 1992
Printed in the USA. All rights reserved.

0893-6080/92 $5.00 + .00

ORIGINAL CONTRIBUTION

Stabilization of Complex Input–Output Functions in Neural Clusters Formed by Synapse Selection

MICHEL KERSZBERG, STANISLAS DEHAENE, AND JEAN-PIERRE CHANGEUX

Institut Pasteur

(*Received* 31 *May* 1990; *revised and accepted* 28 *May* 1991)

Abstract—*We numerically analyze the self-organization of formal neurons disposed in a two-dimensional layer, and receiving inputs from two sets of afferent axons* A *and* B. *The probability for a given afferent to innervate some neuron depends initially on both afferent and target neuron types, which may be excitatory or inhibitory. This early wiring diagram leads to relatively ill-defined functional groups within the neuronal assembly. There follows a period during which the system differentiates, under the presence of external inputs, into groups of neurons with stable input–output relationships. The mechanism proposed for this maturation is based on the management of a limited provision of retrograde trophic factor distributed from postsynaptic neurons to presynaptic terminals whenever a Hebb-like condition is satisfied. Those boutons which do not receive sufficient trophic support ultimately degenerate. The remaining circuitry is characterized by emergent "mexican-hat" type interactions, i.e., short-range excitation vs. longer-range inhibition, and exhibits well-defined functional properties. These final properties are found to depend both on the initial wiring diagram, and on the correlations between the afferent inputs. Thus, increasing the frequency of, e.g., simultaneous activation of* A *and* B, *leads to an increased size of those patches which display activity when* A *and* B *are active together. Patches can be observed which realize in a stable and reliable manner any of the* 16 *Boolean functions of the variables* A *and* B. *Usually, patches endowed with different functions may coexist in a given system.*

Keywords—Neural groups, Neural function, Hebb rule, Synapse elimination.

1. INTRODUCTION

The mechanism by which functional units become established in the peripheral and central nervous systems, in the course of evolution as well as during ontogenesis, is a major issue in the neurosciences (Edelman, 1988; Laufer & Changeux, 1989; Purves, 1989). While the most extreme views have been held in the past on this subject, it has progressively become clearer that the successful buildup of functionally significant neural organizations, in the form of stable "representational patterns" allowing for efficient interactions with the outside world, must result from a complex interplay between genetic and epigenetic phenomena.

In this work, we shall introduce, as a simple, genetically modulated parameter, the *differential probability* that a given neuron class forms synapses on a neuron belonging to another class (Edelman, 1983), while the epigenetic component will be selective, activity-controlled synapse stabilization, known to occur in the peripheral and central nervous systems (Changeux & Danchin, 1976; Purves & Lichtman, 1980). A neuronal connectivity is initially set up in stochastic fashion, taking into account the various probabilities of creating synapses between different pairs of neurons, and between two input fiber bundles and the various neurons in the ensemble; the resulting structure displays rather primitive and unstable functional properties. A plausible mechanism (Changeux & Danchin, 1976) for the subsequent evolution of the system might involve the activity-regulated production and utilization, as a trans-synaptic signal, of trophic factors by the target cells. After a sensitive period, during which certain synapses receive activity-controlled trophic support from their postsynaptic neurons, it is assumed that only those synapses which have accumulated enough trophic factor get stabilized (Alderson, Alterman, Barde, & Lindsay, 1990; Gouzé, Lasry, & Changeux, 1983); we show that, under these conditions, the functional properties of the resulting structure become much stabler than before, and the organism can therefore be said to have developed neural representations ap-

Requests for reprints should be sent to Michel Kerszberg, Neurobiologie Cellulaire, Institut Pasteur, 25, rue du Docteur Roux, F-75724 Paris Cedex 15, France.

propriate to its environment (Kerszberg, 1989). The neuronal population has divided into *patches* exhibiting various functions. The actual functional repartition depends in part on the initial connection probabilities, and, in part, on the actual inputs provided during the stabilization process. When only random inputs are present, the process can be assimilated to a spontaneous maturation of neural circuitry; while in the presence of nonrandom inputs, one can speak of "learning." Learning amounts here to a modification of the relative sizes and locations of functional patches: it can be said to proceed by *selection* among the patches.

In what follows, we use computer simulations of the model to study the influence of selective synapse stabilization on the dynamic formation of functional neuronal clusters.

2. BIOLOGICAL PREMISES

We describe now in more detail the biological premises on which our model rests.

2.1. Differential Afferent Targeting

We shall consider ensembles of neurons on a relatively large scale, where differentiated connectivity patterns are discernible. Thus, we assume that, locally, neural connections are set up in a largely random fashion; while on a global scale, incoming and outgoing connections to a neuronal group are at least partially *structured*, providing a genetic "envelope" to its behavior. In particular, we shall introduce the idea that initial innervation by *afferents* is deterministic to some degree. Of course, such specificity has been established experimentally for a long time and forms the subject matter of neuroanatomy, but it has, until now, seldom been taken into account in model neural networks (Hopfield, 1982; Pearson, Finkel, & Edelman, 1987).

There certainly exists multiple phenomena which may be called upon to account for a differential targeting of neurons by growing neurites, among them morphogenetic gradients (Wolpert, 1969), mechanical, geometrical, and other constraints (Oster & Alberch, 1982), or constraints on the behavior of the growth cone such as differential molecular adhesivity (Edelman, 1988; Harelson, Bastiani, Snow, & Goodman, 1988). Here we shall use as our starting point the fact that excitatory afferents may reach their target system either directly or by way of inhibitory interneurons; the fraction of the total innervation impinging on interneurons is fixed probabilistically. Similarly, both the excitatory afferents and the inhibitory interneurons have different probabilities of forming initial (labile) synaptic connections with excitatory or inhibitory neurons within the group being studied.

2.2. Activity-Dependent Synapse Stabilization

In the adult vertebrate skeletal muscle, each individual muscle fiber is innervated by a single motor axon terminal. Such a pattern develops from a transient *labile* stage, where each endplate receives several motor axon terminals (Bennet & Pettigrew, 1974; Brown, Jansen, & Van Essen, 1976; Redfern, 1970). In the rat, during the first postnatal weeks, such polyneural innervation is altered and supernumerary nerve endings are eliminated (Van Essen, 1982). Activity is required in the neuromuscular system for this phenomenon to take place (Benoît & Changeux, 1975), see also references in (Nelson, Chang Yu, Fields, & Neale, 1989). Similar regressive phenomena have been observed in the central nervous system (Mariani & Changeux, 1980a; Mariani & Changeux, 1980b; Purves & Lichtman, 1980), on a quite comparable quantitative basis in terms of the fraction of synapses finally established on a given neuron—or neurite. As mentioned above, a model has been proposed (Gouzé et al., 1983), which introduces the notion of activity-dependent trophic support of presynaptic terminals by the innervated muscular fibers. Here, we are led to the generalized hypothesis that such retrograde trophic support is also effected by *postsynaptic neurons* in certain areas of the central nervous system (Alderson et al., 1990; Glanzman, Kandel, Schacher, 1989). Accordingly, therefore, survival of initial (labile) connections is controlled by the retrograde release of trophic factor and its utilization by the afferent endings. Stabilization of a presynaptic terminal is thus an epigenetic process dependent on a sufficient level of reinforcement feedback enabled from the postsynaptic end. The latter is assumed to be *activity-dependent* (Cline & Constantine-Patton, 1989; Shatz & Stryker, 1988) and to obey a Hebb-like rule (Hebb, 1949): in the case of *excitatory* termnals, transfer is postulated to occur whenever the pre- and postsynaptic units are simultaneously active; yet, we do not assume the existence of a reverse flow when the Hebb condition is not satisfied. As to *inhibitory* synapses, much less is known experimentally; at this stage of knowledge, we shall suppose that retrograde transport of trophic factor takes place whenever the *postsynaptic* neuron *is activated*, while the inhibitory terminal is *not* firing (Rauschecker & Singer, 1981). There are of course other logical possibilities, but we have not explored their consequences here.

3. PRINCIPLES OF THE SIMULATIONS

In this section, we should like to introduce in some detail the system we simulate. We describe first the functioning of the basic computational units, their organization into a connected set, and the way in which inputs are fed into this network. We then de-

scribe our model of selective synapse stabilization and elimination, the main epigenetic process which we postulate here.

3.1. Formal Neurons

We use McCulloch–Pitts neurons (McCulloch & Pitts, 1943). It is assumed that each neuron's axonal processes are characterized by the *same* neurotransmitter whose type does not change in the course of time (Dale's law). Thus, neurons are either excitatory or inhibitory, but not both. Note that there is evidence for variations in neurotransmitter type during ontogenesis (Brecha, Johnson, Peiche, & Wässle, 1988; Schotzinger & Landis, 1988), as well as for the importance of the receptor in determining the sign of synaptic interactions, (Kandel, 1976) but we shall not take such effects into account here.

The operation performed by each formal neuron consists of:

- summing its inputs, weighted by a weight W_{ij}, where i denotes the postsynaptic, and j the presynaptic neuron, respectively;
- taking, on the basis of the above sum, the decision to generate or not an action potential. We assume here that the decision is a *deterministic* one: if the sum is greater than some threshold value T_i, firing takes place, otherwise not. Although this is not essential, and influences the results only in the details of the final wiring, it was also assumed, for computational convenience, that T_i is itself adjustable as a function of global excitation vs. inhibition afferent to the neuron: more precisely, we take

$$T_i = T_0 + T_+ \sum_{W_{ij}>0} W_{ij}, \qquad (1)$$

where $T_0 = 7.00$ and $T_+ = 0.20$. (Thus, at the neuromuscular junction, e.g., activity is, to some extent, self-regulated. Partial denervation, for example, will lead to an increase in sensitivity on the part of the postsynaptic end plate (Pumplin & Fambrough, 1982). Some regulation of the activity is certainly to be expected in the central nervous system as well (Bear, Cooper, & Ebner, 1987).)

3.2. Differential Targeting: The Network and its Inputs

We simulate a group G consisting in an equal number of excitatory and inhibitory neurons, arranged randomly on a periodic, 2-dimensional (30×30) array, and connected in a network of short-range interactions. Each neuron forms an average of $n = 30$ connections with neighboring neurons. The probability that a connection is established initially (i.e., before the stabilization process takes place) decreases exponentially with the interneuronal distance: we shall take a decay length of $l = 1$, *independent* of neuron type. The probability that a synapse is formed is also taken to be independent of the pre- and postsynaptic neurons' types.

In addition to the connections from its neighbors inside the group G, each neuron also receives m *inputs*; we take $m = 10$ on the average. Each input may originate from either one of two afferents A and B. Two sorts of parameters related to the afferent A- and B-synapses are used to introduce the notion of differential adhesivity between neuronal processes (see Figure 1). They do so in a complementary way. First, terminals from different origins may form on excitatory or inhibitory neurons in G with unequal probabilities, namely the probability that an A-synapse targets an excitatory neuron is taken to be $p_{A,E}$ and may in general have a value different from 1/2, or from $p_{B,E}$, the corresponding probability for a B-synapse. Furthermore, we assume that axons originating from A are excitatory with a probability p_A^{exci}, while the corresponding probability for B-synapses, p_B^{exci}, may sometimes be different from that. One can interpret these various probabilities in several biologically plausible ways, as seen in Section 2.1; thus, the differential targeting of excitatory and inhibitory neurons by axons of different origins might, for instance, be caused directly by a differential adhesivity as may originate from the presence of differing molecular labels on the surface of the concerned cells (Edelman, 1983). That afferent fibers have an excitatory *or* inhibitory effect could, e.g., be the result of a wiring diagram in which all afferent

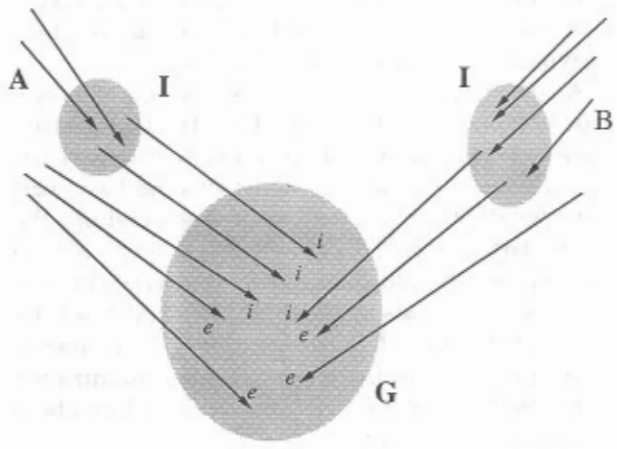


FIGURE 1. Example for the differential targeting of neural groups by afferents originating, e.g., form two separate sensory modalities. Axons from, say, *A*, may form synapses either on *G*-neurons directly by way of inhibitory interneurons (groups labelled *I*) (probability p_A^{inhi}); in addition, afferents originating directly or indirectly from *A* may target excitatory or inhibitory neurons in *G* (probabilities $p_{A,E}$ and $p_{A,I}$, respectively).

fibers are excitatory, but where a variable proportion of them target *inhibitory interneurons* in groups I (see Figure 1), which in turn form synapses with the neurons in G (and essentially none between themselves).

All inputs of type A or of type B are activated simultaneously, indicating the presence of a stimulus on the corresponding channel. This "external stimulation" is not necessarily structured, i.e., it may well be that stimuli A and B occur in a stochastically independent fashion. We shall see however that it may be very important and interesting to study the effect of *correlations* between stimuli; we shall therefore assume in general three independent probabilities p_{AB}, $p_{A.}$, $p_{.B}$ for the three events:

- A and B occur simultaneously at some time,
- A occurs alone,
- B occurs alone.

Of course the sum $S = p_{AB} + p_{A.} + p_{.B}$ is less than or equal to 1, and the difference $1 - S$ is the probability that no stimulation takes place.

3.3. Selective Synapse Stabilization

The course of the simulations is the following. The network is kept working for a certain number of time units t; typically $t \simeq 2{,}500$. At each time step, *all* neurons read their inputs, both internal and external, and update their own state accordingly (see Section 3.1). The external stimuli are adjusted every five time steps: we have checked that this interval of $t_s = 5$ time units is usually sufficient for the network to reach an essentially stationary state. At this stage, all neurons also provide their presynaptic terminals with trophic factor. Note that, although we take t_s such that stationarity is reached, this may not be a prerequisite for satisfactory performance.

As discussed in Section 2.2, we assume that synaptic boutons are initially established in a *labile* state. They may exist in two additional states, namely *stabilized* and *degenerated*. A labile synapse has a certain probability per unit time of degenerating. We set the latter to be 1/2,500, so that after $t \simeq 2{,}500$ the synapses still alive are mostly stabilized (the importance of this assumption for the model will be discussed in some detail in Section 4.1.1). Stabilization is dependent on the presynaptic concentration of trophic factor in a given terminal, σ_{ij}, which obeys the following equations:

$$\Delta\mu_i^s = -\delta_{\text{Hebb}} k_0 \mu_i^s, \tag{2}$$

$$\Delta\sigma_{ij} = -\Delta\mu_i^s. \tag{3}$$

These equations express the retrograde diffusion (k_0 is the diffusion constant) of trophic factors μ_i^s, of type s ($s = I$ or E for inhibitory or excitatory) and present in *postsynaptic* neuron i. This diffusion away from neuron i [whence the minus sign in eqn (2)] is triggered whenever the Hebb-like rule discussed above is satisfied ($\delta_{\text{Hebb}} = 1$, otherwise $\delta_{\text{Hebb}} = 0$). In the presynaptic terminal, the factor is converted to σ form [eqn (3)]. The trophic support of excitatory and inhibitory synapses is thus assumed to depend on two different agents, μ^E and μ^I. The differences in eqns (2) and (3) are taken per time step. Initially, $\sigma_{ij} = 0$. A synapse is assumed to stabilize as soon as $\sigma = 1$. We take $k_0 = 0.01$, and adjust the initial values of μ^s so that stabilization of strongly reinforced synapses is almost certain within a time $t \simeq 2{,}500$. Usually, the μ^s are initially set around 8 to 20. With these values, typically, the fraction of surviving synapses is, after stabilization, on the order of 1/5 to 1/8: this is in rough agreement with neuroanatomical data (Purves, 1989). In order that the global strength of synaptic interactions be maintained approximately constant throughout the stabilization process, we introduce the additional assumption (Gouzé et al., 1983) that the strength of a synapse depends on the amount of accumulated trophic factor. Thus

$$W_{ij} = a_0 + b_0\sigma_{ij}, \tag{4}$$

where a_0, b_0 are constants (see below), and σ_{ij} is the presynaptic concentration of trophic factor in the terminal. In view of the usual values for the elimination ratio, we choose $a_0 = 1$, $b_0 = 4$, so that if $W_{ij} = 1$ initially, it will reach 5 after stabilization.

4. MATURATION OF NEURAL CONNECTIONS

We study now the way in which afferent input activity may regulate the various functional groups formed by synaptic elimination. The short range of the connections allows a rich situation where neurons are self-organized in multiple, distinct functional patches.

4.1. Coexistence of Groups With Differing Functionality

In all of the following simulations, the initial *internal* connectivity is the same; therefore, all the examples of stabilized structures or functional maps which we display may be compared with one another in order to gauge the relative importance of various factors in the final establishment of neural patterns.

4.1.1. *An Undifferentiated Architecture.* We first look at the case of an assembly built under totally bias-free conditions, i.e., $p_A^{\text{exci}} = p_B^{\text{exci}} = 0.5$, $p_{AB} = p_{A.} = p_B = 0.25$. Figure 2 (parts *a* and *c*, *d*) displays a typical network configuration after synaptic elimination, and the functional behavior before and after it. We do not show here the connection diagram

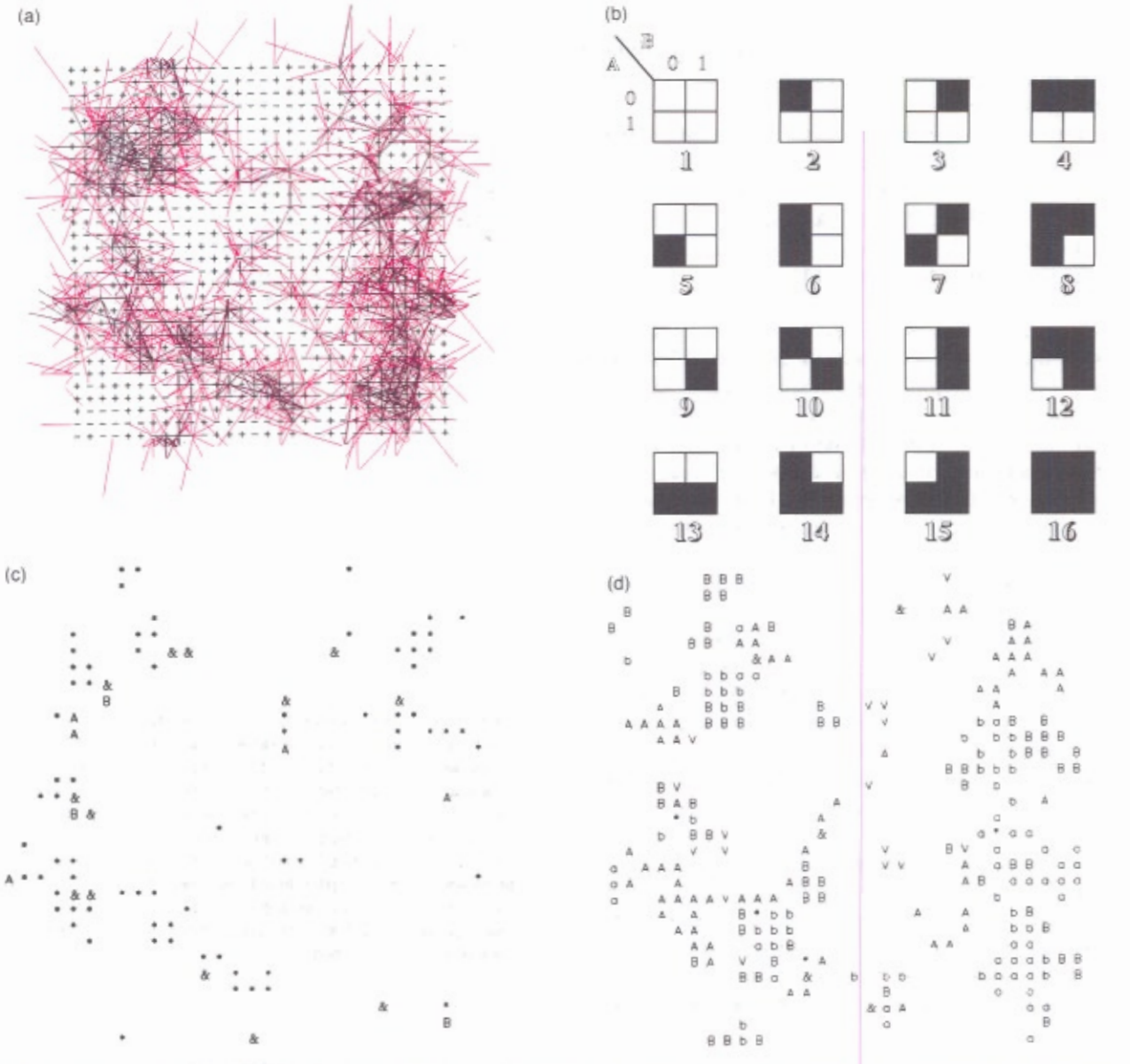


FIGURE 2. Selective synapse stabilization in a population with little preexisting organization: afferents axons target neurons in *G* or in *I* with equal probabilities, i.e., $p_A^{exci} = p_B^{exci} = 0.5$. The input environment is also minimally structured, all stimulus combinations being equally probable: $p_{AB} = p_{A\bar{B}} = p_{\bar{A}B} = 0.25$. Black denotes excitatory, red inhibitory connections or neurons. a) The stabilized connections. Several groups can be identified, each with an excitatory core surrounded by an inhibitory shell. b) The sixteen Boolean functions of two variables *A* and *B*. Each function is specified by the values it takes for the four combinations of (*A*, *B*); the values are symbolized here by white (0, not firing) or black (1, firing). Usual names for the functions are: 1: *FALSE*; 2: $A \bar{\vee} B$; 3: $B\&\bar{A}$, also denoted as *b*; 4: $\bar{A}$; 5: $A\&\bar{B}$, also denoted as *a*; 6: $\bar{B}$; 7: A X B [exclusive-*OR*]; 8: $\overline{A\&B}$; 9: $A\&B$; 10: $A \Leftrightarrow B$; 11: B; 12: $A \Rightarrow B$; 13: *A*; 14: $A \Leftarrow B$; 15: $A \vee B$ [inclusive-*OR*]; 16: *TRUE*. To this, we add a 17th category of neurons, denoted by a star ∗, whose activity after the system has stabilized depends not only on the inputs A and B, but also on the initial state of the system itself. c, d) Summary of the increased and improved differentiation in functionality brought about through selective synapse stabilization and elimination. Boolean function of the inputs, as implemented by each neuron, are displayed before stabilization (c) and after (d); in c, most neurons are activated in erratical fashion (∗), while in d, various functional patches are discernible, each realizing a given Boolean function of *A* and *B*.

before elimination, as this would consist essentially in an apparently structureless jumble of connections. Because of random fluctuations in the way connections are established, however, neural *activities* need not be totally uncorrelated in such a system.

After a period of activity, during which selective synapse stabilization is allowed to take place, the connections have shrunk to those on Figure 2a. In this figure, black color codes for excitatory, and red for inhibitory connections. Each site on the picture corresponds to one neuron in the array. We see that neuron clusters have been defined, grouping those

which had a slightly stronger initial functional coupling. These are held in simultaneous activity through excitatory links. Between those "cores," inhibitory "bridges" can be seen.

Let us examine the statistical length properties of our system's final connectivity. The existence of clusters should lead to an effective interaction among neurons which is of the well-known mexican-hat form (Linsker, 1986), namely short-range excitation and long-range inhibition. In addition, however, but partially related to this, the length distribution of inhibitory connections is skewed towards long distances as compared to excitatory connections. This can be seen on connection length histograms, as displayed on Figure 3. It is important to note that these mexican-hat-type features of the effective interneuronal interaction are not an assumption of the model, but rather a *result* which appeas to hold in a variety of situations. This outcome can be understood rather simply in intuitive terms. Clearly those neurons that have a tendency to be active simultaneously initially tend to form local groups, due either to their interconnections happening to be mostly excitatory, or to the chance existence of several similar afferent connections targeting a given neighborhood. The excitatory connections within groups will be reinforced, leading to further growth; while since the groups will sometimes display nonsimultaneous activity, inhibitory intersynapses between groups will likewise receive trophic factor.

In this context, it is interesting to follow the time course of the elimination and stabilization process. We have plotted on Figure 4 the fractions of stable and eliminated synapses as a function of time, as well as the proportions of trophic factor still available to neurons for synapse support. We see (a) that elimination starts very smoothly, and proceeds faster as the mean labile synapse lifetime is approached. This might be an artifact of the model and can be explained in the following way. In our computations, the retrograde transfer of trophic factor from postsynaptic neurons to presynaptic terminals is a diffusive process (see eqn (2)) and therefore the concentration of trophic factor cannot decrease below a certain level; as can be seen on Figure 4b, the fractions of available trophic factor do indeed remain high, and it is therefore *not* the depletion of these factors which can be responsible for "starvation" and death of the terminals (see Catsicas & Clarke, 1987; Lowrie, Krishnan, & Vrbová, 1982 for some evidence in favor of this). Rather, the most eligible terminals collect their due of trophic factor and stabilize, while less eligible terminals remain in a labile state and therefore die at a constant rate, this leading finally to their total disappearance. In a model with nondiffusive transfer of trophic factor, or

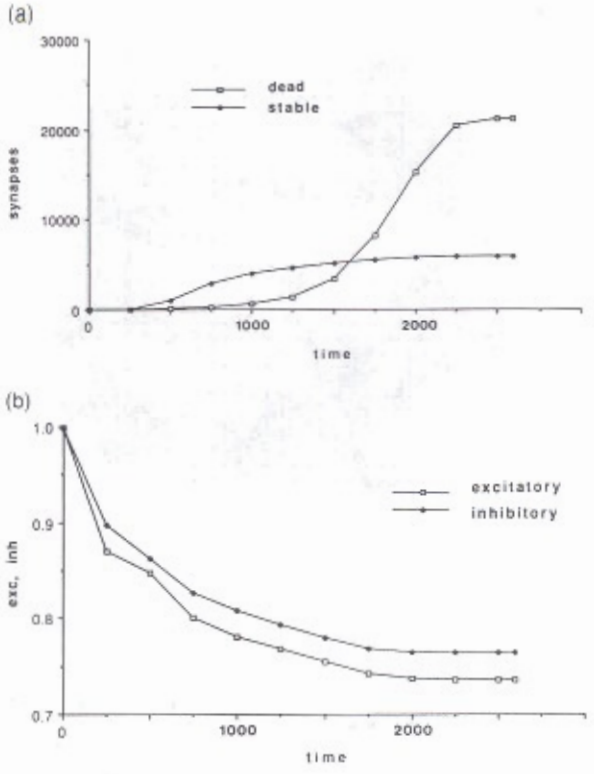


FIGURE 4. Time course of synapse stabilization in the simulations of Figure 2. (a) Number of stabilized and eliminated synapses as a function of time. Note that the stabilization process saturates around t = 2,500; afterwards, synapses that are still labile are eliminated massively. (b) Postsynaptic neurons are assumed to have available, from the outset, a fixed amount of trophic factor to be delivered to excitatory or inhibitory presynaptic terminals. Here, the fraction of trophic factor still not delivered is plotted as a function of time. We see that, by t = 2,500, retrograde transfer of the trophic factors has almost ended.

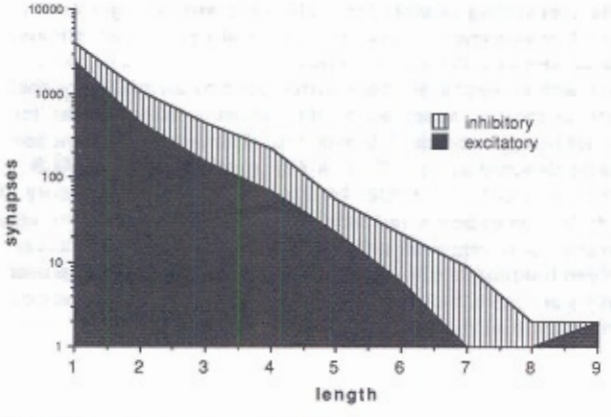
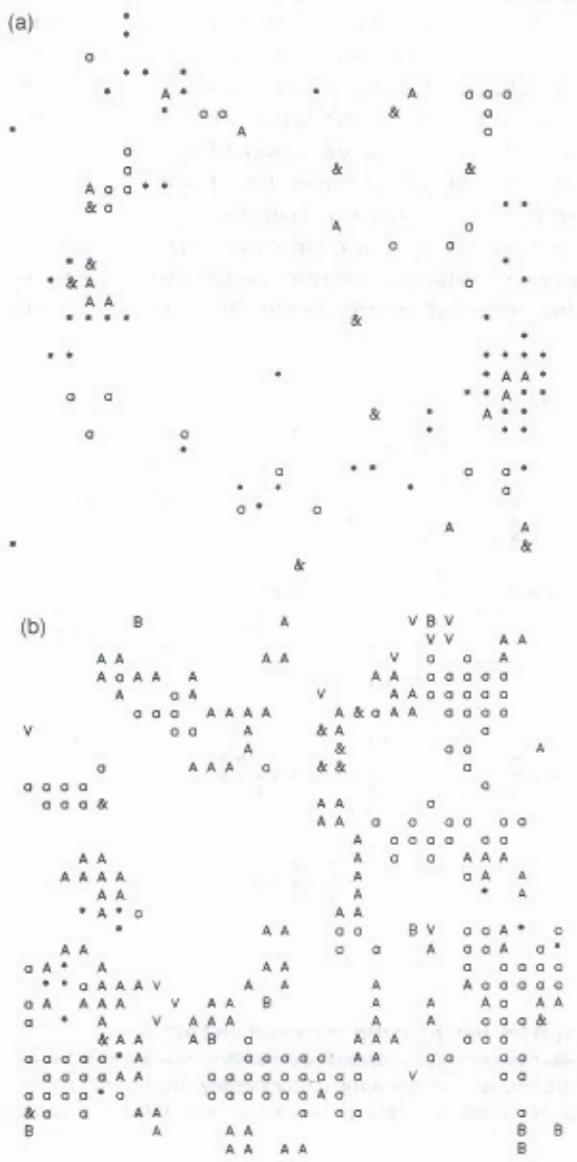


FIGURE 3. Number of stabilized synapses at the end of the simulation on Figure 2, plotted on a logarithmic scale as a function of synapse length. Shaded: number of excitatory synapses; hatched: excess of inhibitory over excitatory synapses. The excess is partly responsible for a "mexican-hat" type of interaction between neurons: excitatory at short range, inhibitory at longer distances (see also Figure 2a).

without spontaneous degeneration of labile synapses, stabilization would follow a rather different time evolution. However, no drastic modifications should be expected, at least as far as the end result is concerned.

The functional behavior of a given neuron is described in terms of the sixteen possible Boolean functions of two Boolean variables. Thus, if a neuron fires reproducibly (i.e., with good independence from the initial network's conditions) whenever input A is active while B is not, we say that the neuron implements the Boolean function "$A\&\overline{B}$", or a, which is true *if and only if* A is true *and* B is not. The sixteen Boolean functions of two variables are summarized in Figure 2b ("a" corresponds there to function 5). It is striking to notice, on Figure 2c and d how neural activity, at first fleeting and ill-determined, becomes, after synapse elimination, much more clearly established in terms of independence from the initial state of the neuron array (after at time $t = t_s$), and how functional groups have turned reproducible in their activity. Before stabilization, only 21 neurons display any interesting functional properties, while after it, there exists 321 such units. The mature functionalities displayed are not random; certain functions are conspicuously rare (but not always entirely absent), such as the "exclusive-or" (true if A is active or B is active but not both). Such unequal repartition can be expected on very general grounds (Mukamel & Kerszberg, 1989); but since the system is built symmetrically with respect to A or B, no bias in favor of, say, function "a" against function "$B\&\overline{A}$" (i.e., "b") can be noticed.

4.1.2. *Influence of Architectural Preorganization.* The simulations summarized on Figure 5 pertain to a case of initial structural bias, namely we now have $p_A^{\text{cxci}} = 0.6$, $p_B^{\text{cxci}} = 0.4$. It can be seen very clearly on part



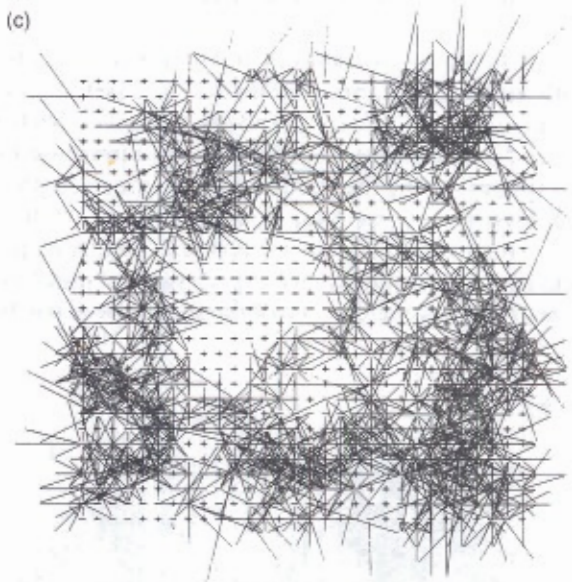


FIGURE 5. Functionality and architecture in a system with more initial structure than in Figure 2: here $p_A^{\text{cxci}} = 0.6$, $p_B^{\text{cxci}} = 0.4$. On the other hand, the inputs are still random. (a) Functionality before synapse stabilization: note that, although still erratic to a large extent, function is more clearly defined than in 2c; a bias appears in favor of those functions, such as *A* or a, for which *A* and *B* must play excitatory and inhibitory roles, respectively. (b) Improved functionality after synapse elimination. Large and stable clusters appear, displaying mostly the functionalities "favored" by the afferent fiber connection bias. Comparing with 2d, we see that certain clusters are present in both cases, but with a different activity pattern; while others have appeared or disappeared. Here, therefore, we observe more than just an enhancement of existing connectivity groups: there is an actual *selection* and *alteration* of groups according to the synchronized activities of their members. (c) The same remark holds true for the connectivity network as well: it is interesting to compare this picture with 2a, and watch how the groups have been sculpted neuroanatomically by the initial connectivity bias.

b of the figure how this has altered the final, stabilized functional behavior map in the direction of increased contingents of neurons being enrolled by *A*-excited, *B*-inhibited functions such as "*A*" and "a"; the internal connection map has also changed to quite an extent, and although certain groups can still be identified on both Figs. 2a and 5c, many differences are also visible.

One may say that the effect of selective elimination is the amplification, and the expression at the level of connection clusters, of functional correlations preexistent in the system because of its initially random setup. We have just demonstrated now that introducing such correlations at a level well above that of chance fluctuations (see Section 4.1.1), as modulated through variations in the probabilities for *A*- or *B*- afferent fibers to target different cell types within the system (G) or interneurons (I), i.e., by changes in the parameters p_A^{exci} and p_B^{exci}, leads to a remarkable alteration in the structure and functioning of the final, mature neural system.

4.1.3. *Influence of Biases in Input Stimuli.* We presently simulate synaptic elimination in a system completely identical (initially) to that studied in Section 4.1.1. Now, however, the input stimuli presented to the system during the stabilization/elimination phase are *biased*, i.e., we take $p_{AB} = 0.30$, $p_{A.} = 0.10$, $p_{.B} = 0.60$. Thus, stimulus *B* appears *alone* sixty percent of the time: clearly we expect such an input bias to lead to a degenerescence of those functional patches in the system which were devoted to logical function dependent on the presence of input *A*, and we now show (see Figure 6) that this is indeed the case. Comparison of Figs. 2d and 6b is quite convincing in this respect. Note also that the connectivity structure has changed rather conspicuously (see Figs. 2a and 6a). We have seen in the previous subsection how preexisting structural biases may have rather similar influences on the final wiring; thus, one can surmise that situations may exist where competition between the two arises: the next Section is devoted to just such a case.

4.2. Interaction Between Internal Architecture and External Stimulation

In the following computation, we introduce both a structural bias in favor of excitatory *A*-connections (see Section 4.1.1) and a stimulational bias favoring *B*-inputs. The results are displayed on Figure 7. The pictures shown there should be compared with those of Figs. 5 and 6. We see that the system has tried to find a "compromise," so to speak, between the somewhat conflicting requirements arising from both biases: thus, a large patch has arisen where the functional behavior is actually according to the "&" logical function, which treats the *A* and *B* inputs in a completely *symmetrical* fashion!

Of course, just as *competition* may exist between biases of differing origins, *cooperation* is also possible, although its effects are, by necessity, less spec-

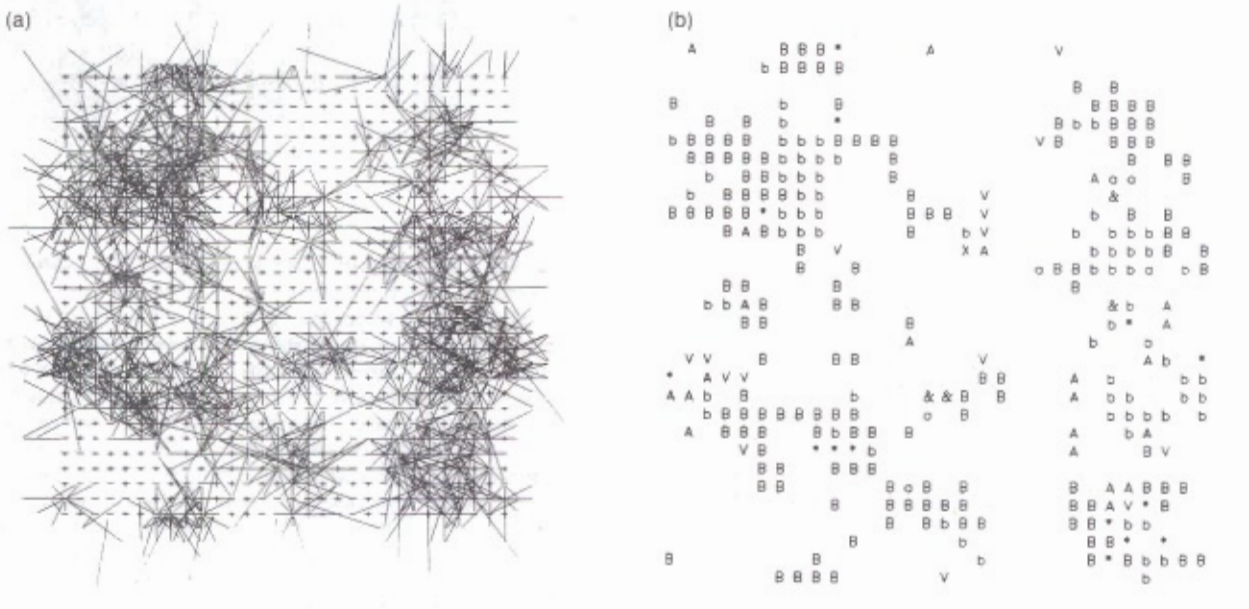


FIGURE 6. Structure and function after synapse stabilization in a system with no initial organization ($p_A^{\text{exci}} = p_B^{\text{exci}} = 0.5$) but whose inputs were correlated: $p_{AB} = 0.30$, $p_{A.} = 0.10$, $p_{.B} = 0.6$. (a) Neuroanatomical connection pattern (compare Figs. 2a and 5c!). Correlations among inputs are clearly capable of modulating the group configuration quite deeply. (b) Functionality: the dominant groups now exhibit Boolean functions for which *B* is most excitatory, while *A* is indifferent (compare Figure 2d).

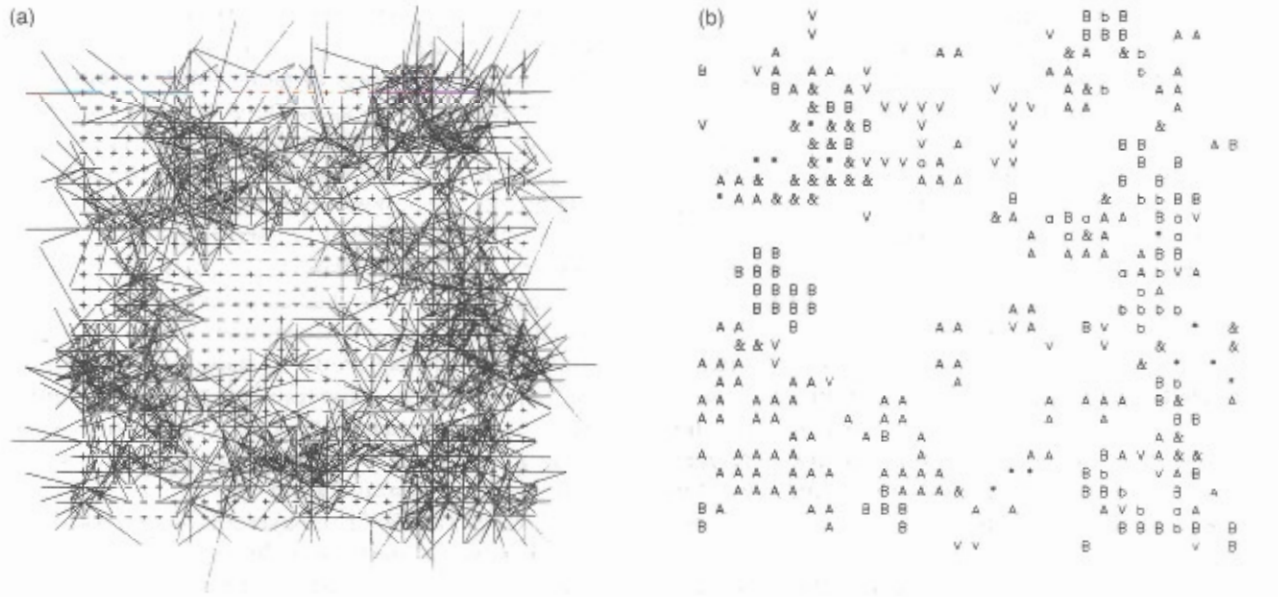


FIGURE 7. A mature system initially set up with the architectural bias corresponding to that in Figure 5, and which has been subjected to biased inputs just like those of Figure 6. The results of the competition between endogenous and exogenous factors are clearly observable. (a) Connectivity; (b) Functionality. This is much more varied than in the previous cases, with the appearance of stably operating patches which implement functions *V* or &.

tacular. The point we want to stress is that the final connectivity and functionality are the result of the combined effects of initial architecture, modulated to some extent by externally imposed stimulation.

In particular, it is important to realize that there may exist several ways to obtain a circuit exhibiting a particular behavior. However, although the final wiring may have similar properties, the evolutionary value of the various routes to this wiring diagram may be very different. Systems whose function has been established early in phylogenetic history are likely to have their synaptic contacts established in a way which is relatively independent from the environment; while connections whose importance is more recent, or which must be set up in a strong environmentally adaptive fashion, are more likely to be not so well defined in terms of their initial architecture (Cairns, Gariépy, & Hood, 1990; Kerszberg, 1989).

5. OUTLOOK

We have analyzed numerically a model for the self-organization of neural assemblies into clusters of cells with characteristic, input-related response. These clusters appear as a global ordering which stems from a nonlinear interaction among local rules for synapse and neuron operation and stabilization. A double set of influences shapes these neural ensembles: those pertaining to the (genetically determined) way in which initial incoming connections from two afferences A and B are set up in a partly stochastic fashion, on the one hand (Kerszberg, 1989); on the other, the existence of environmentally determined statistical correlations among the occurrences of external stimuli A and B. Function, defined here as a precise relationship between stimulation—on channels A, B—and cell response, will at first appear rather fleeting and unstable, although not completely absent. Its full development will be the result of a maturation process. The latter is modelled within the general framework of selectionist schemes (Dawkins, 1971; Changeux & Danchin, 1976; Edelman, 1987) and results in the differential stabilization of a particular subset of connections. The stabilization mechanism proposed here operates through the management of a limited stock of trophic factor, distributed from postsynaptic neurons to afferent terminals under an activity-dependent Hebb-like rule (Gouzé et al., 1983). Those terminals which do not accumulate adequate trophic support degenerate. The development of fully stable functional sets of neurons is therefore partially regulated by neural activity, i.e., by external (A, B) stimulation patterns; such regulation is effective during the course of a *sensitive* period (Marler & Terrace, 1984) linked here to the particular time at which the selective stabilization of synapses, or their repression, is to occur.

A variety of endogenous and exogenous factors thus participate in the morphogenenesis of topographically and functionally stable neural groups. The endogenous factors include the differentiated probabilities for the establishment of terminals between various neural classes, and the activity-dependent re-

inforcement and ultimate stabilization of some terminals. Neural clusters are shaped by, and selected through the imposition of input patterns which indirectly control the activities of the various groups. The main novelty of our work has been the combination of these elements and their incorporation into a space-dependent model for the appearance of functional response properties in the nervous system. While the system is, initially, only weakly differentiated in terms of both connectivities and functional specialization, progressive refinement occurs throughout the selective stabilization process. Modifying the relative frequencies of various stimulus combinations leads to the selection (i.e., growth and performance improvement) of the neural clusters detecting precisely those combinations or related ones. Normally, clusters endowed with different response characteristics will coexist within the system. These clusters usually consist of neurons which excite each other mutually, and inhibitory "bridges" may be observed between the clusters. This leads to a network where excitation emerges as a short-range property, and inhibition as a longer range one. In the models we discuss below, and this is also the case for previous work, for example, that of (Willshaw & von der Malsburg, 1976) on the self-organized appearance of column-like structures, it must be noted that such important features as short-range excitation vs. long-range inhibition are introduced at the outset, while here they are a result of the system's operation.

Fine details of how trophic factor is retrogradely transferred at each synapse will, in general, have an influence on the final structure and its performance. The further study of such transfer will thus be an important research objective. Similarly, the final state of the system must depend on the regulation of overall activity levels, both at the cellular scale and on a more global one, through, for example, the action of various diffusible factors.

Much work has been devoted lately to certain aspects of the problem we have considered; thus, some comparisons are in order. One of the major points here is that we have tried to concentrate on questions of principle, seeking to define the *simplest*, most elementary model capable of acquiring the properties sought for. Thus, no detailed account is taken of synaptic transmission, electrochemical characteristics of the neuron membrane, impulse generation, etc. This is in contrast with the work reported in (Pearson et al., 1987). These authors undertake to study the modulations of finger dominance areas in somatosensory cortex as resulting from sensory deprivation experiments (Merzenich, Recanzone, Jenkins, Allard, & Nudo, 1988): in a language closer to the one used here, we might say that their neuronal groups' logical function is to be active when and only when the signal corresponding to that finger is present. We, in contrast, study only the case of *two* input variables explicitly, but we consider as our foremost objective the detection of much subtler correlations among these. We do so in the context of *epigenesis*, and this is also a major additional difference.

A remarkably detailed analytical treatment of ocular-dominance column development has been proposed by (Miller, Keller, & Stryker, 1989). Various remarks previously made apply also to this work. The specific response studied is limited, in this case, to ocular dominance. On the other hand, the authors proceed by solving global dynamical equations, in contrast to our microscopic approach. The results are therefore exact (within the framework of the model) and afford in-depth comparison with a well-studied experimental system.

In (Linsker, 1986), a layered, feedforward-type neural network model was developed. The appearance of various cell types exhibiting specialized responses was studied. The dynamics of the system was not simulated directly but examined rather through the analysis of more global equations. Probably the most important difference with the present work is that Linsker did not allow for "reentrant" connections. We have demonstrated here how the detection of complex correlations among inputs is possible in a monolayer, reentrant organization, and how this capability is being modulated by various means.

The acquisition of skills in a way linked to the preferred occurrence of specific stimulation may be considered as a form of *supervised* learning, the supervision being provided by a nonrandom environment; when skills are acquired without such occurrence, one may speak of *unsupervised* learning: the supervision must then have been provided through the selective effects exercised by past environment on the species. One could say that our model articulates together these two types of learning phenomena.

REFERENCES

Alderson, R. F., Alterman, A. L., Barde, Y.-A., & Lindsay, R. M. (1990). Brain-derived neurotrophic factor increases survival and differentiated functions of rat septal cholinergic neurons in culture. *Neuron*, **5**, 297–306.

Bear, M. F., Cooper, L. N., & Ebner, F. F. (1987). A physiological basis for a theory of synapse modification. *Science*, **237**, 42–48.

Bennett, M. R., & Pettigrew, A. G. (1974). The formation of synapses in striated muscle during development. Journal of Physiology (London), **241**, 515–545.

Benoit, P., & Changeux, J. P. (1975). Consequences of tenotomy on the evolution of multiinnervtion in developing rat soleus muscle. *Brain Research*, **99**, 354–358.

Brecha, N., Johnson, D., Peichl, L., & Wässle, H. (1988). Cholinergic amacrine cells of the rabbit retina contain glutamate decarboxylase and γ-aminobutyrate immunoreactivity. *Proceedings of the National Academy of Sciences* (USA), **85**, 6187–6191.

Brown, M. C., Jansen, J. K. S., & Van Essen, D. C. (1976). Polyneural innervation of skeletal muscle in newborn rats and its elimination during maturation. *Journal of Physiology* (London), **261**, 387–422.

Cairns, R. B., Gariépy, J. L., & Hood, K. E. (1990). Development, microevolution and social behavior. *Psychological Reviews*, **97**, 49–65.

Catsicas, S., & Clarke, P. G. H., (1987). Abrupt loss of dependence of retinopetal neurons on their target cells, as shown by intraocular injections of kainate in chick embryos. *Journal of Comparative Neurology*, **262**, 523–534.

Changeux, J. P., & Danchin, A. (1976). Selective stabilisation of developing synapses as a mechanism for the specification of neuronal networks. *Nature*, **264**, 705–712.

Cline, H. T., & Constantine-Patton, M. (1989). NMDA receptor antagonists disrupt the retinotectal map. *Neuron*, **3**, 413–426.

Dawkins, R. (1971). Selective neurone death as a possible memory mechanism. *Nature*, **229**, 118–119.

Edelman, G. M. (1983). Cell adhesion molecules. *Science*, **219**, 450–457.

Edelman, G. M. (1987). *Neural darwinism*. New York: Basic Books.

Edelman, G. M. (1988). *Topobiology*. New York: Basic Books.

Glanzman, D. L., Kandel, E. R., & Schacher, S. (1989). Identified target motor neuron regulates neurite outgrowth and synapse formation of aplysia sensory neurons *in vitro*. *Neuron*, **3**, 441–450.

Gouzé, J. L., Lasry, J. M., & Changeux, J. P. (1983). Selective stabilization of muscle innervation during development: A mathematical model. Biological Cybernetics, **46**, 207–215.

Harelson, A. H., Bastiani, M. J., Snow, P., & Goodman, C. S. (1988). From cell ablation experiments to surface glycoproteins: Selective fasciculation and the search for axonal recognition molecules. In S. S. Easter, Jr., K. F. Barald, & B. M. Carlson (Eds.), *From message to mind* (pp. 96–109). Sinauer Associates.

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

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* (USA), **79**, 2554–2558.

Kandel, E. R. (1976). *Cellular basis of behavior*. San Francisco: W. H. Freeman.

Kerszberg, M. (1989). Genetics and epigenetics of neural function: A model. *Journal of Cognitive Neuroscience*, **2**, 51–57.

Laufer, R., & Changeux, J. P. (1989). Activity-dependent regulation of gene expression in muscle and neuronal cells. *Molecular Neurobiology*, **3**, 1–35.

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

Lowrie, M. B., Krishnan, S., & Vrbová, G. (1987). Permanent changes in muscle and motoneuron induced by nerve injury during a critical period of development of the rat. *Developmental Brain Research*, **31**, 91–101.

Mariani, J., & Changeux, J. P. (1980a). Multiple innervation of Purkinje cells by climbing fibers in the cerebellum of the adult staggerer mutant mouse. *Journal of Neurobiology*, **11**, 41–50.

Mariani, J., & Changeux, J. P. (1980b). Etude par enregistrements intracellulaires de l'innervation multiple des cellules de purkinje par les fibres grimpantes dans le cervelet du rat en développement. *Computer Rendus Academie des Sciences (D) (Paris)*, **291**, 97–100.

Marler, P., & Terrace, H. S. (1984). *The biology of learning* (Berlin, Heidelberg: Springer Verlag.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysiology*, **5**, 115–133.

Merzenich, M. M., Recanzone, G., Jenkins, W. M., Allard, T. T. & Nudo, R. J. (1988). Cortical representational plasticity. In P. Rakic & W. Singer (Eds.), *Neurobiology of neocortex*. Chichester: Wiley-Interscience.

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

Mukamel, D., & Kerszberg, M. (1989). *unpublished*.

Nelson, P. G., Yu, Chang, Fields, R. D., & Neale, E. A. (1989). Synaptic connections *in vitro*: Modulation of number and efficacy by electrical activity. *Science*, **244**, 585–587.

Oster, G. F., & Alberch, P. (1982). Evolution and bifurcation of developmental programs. *Evolution*, **36**, 444–459.

Pearson, J. C., Finkel, L. H., & Edelman, G. M. (1987). Plasticity in the organization of adult cerebral cortical maps: A computer simulation based on neuronal group selection. *Journal of Neuroscience*, **7**, 4209–4223.

Pumplin, D. W., & Fambrough, D. M. (1982). Turnover of acetylcholine receptors in skeletal muscle. *Annual Review of Physiology*, **44**, 319–335.

Purves, D. (1989). *Body and brain*. Cambridge, MA: Harvard University Press.

Purves, D., & Lichtman, J. (1980). Elimination of synapses in the developing nervous system. *Science*, **210**, 153–157.

Rauschecker, J. P., & Singer, W. (1988). The effects of early visual experience on the cat's visual cortex and their possible explanation by Hebb synapses. *Journal of Physiology*, **310**, 215–239.

Redfern, P. A. (1970). Neuromuscular transmission in newborn rats. *Journal of Physiology* (London), **209**, 701–709.

Schotzinger, R. J., & Landis, S. C. (1988). Cholinergic phenotype developed by noradrenergic sympathetic neurons after innervation of a novel cholinergic target *in vivo*. *Nature*, **335**, 637–639.

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

Van Essen, D. C. (1982). Neuromuscular synapse elimination: Structural, functional and mechanistic aspects. In N. C. Spitzer (Ed.), *Neuronal development*. New York: Plenum Press.

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

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology*, **25**, 1–47.