Nervous Structures with Dynamical Links

C. von der Malsburg
Max-Planck-Institut für Biophysikalische Chemie,
D-3400 Göttingen, Federal Republic of Germany

Brain Theory / Computer Experiments / Dynamic Networks
Nervous System / Non-linear Phenomena

Connectivity patterns (instead of mere activity distributions) are argued to be the natural objects of fast brain dynamics. They are to be understood as a basis for self-organization. Natural assumptions on cooperation and competition lead to a conjecture regarding the type of dynamically preferred connectivity patterns. The conjecture is supported and illustrated by dynamic simulation. Applications to sensory segmentation, to invariant pattern recognition, and to resonant projection between homeomorphic structures are discussed.

Introduction

According to the current neurobiological picture of the brain, the essential elements and variables for its function are the following. Nerve cells (neurons) send out signals along fibres (axons). Signals have the form of unitary electrochemical pulses (action potentials). Signals are transferred from axons to dendritic and somatic membrane surfaces of the receiving cells by specialized organelles (synapses). Signals may have an excitatory or an inhibitory influence (postsynaptic potential) on the membrane potential of the receiving cell (in addition to biochemical effects). The integration of different influences is non-linear (details, however, are unknown, and models often assume linear summation). The resulting signal, if exceeding a threshold, controls the rate of output impulses. There are several thousand million nerve cells in the human cerebrum, and there are between a thousand and ten thousand synapses on a typical large neuron.

Nervous signals are to interpreted (according to current views) in a mean frequency code. The resultant signal varies on the time-scale typical of global processes in the brain, tenths of a second. Over short times, the network struc-
ture of the brain is commonly regarded as fixed. (It is supposed to be changed slowly by synaptic plasticity, which is, however, left out of consideration in this paper). According to current views, the state of the brain in a given moment is therefore completely described by the vector of firing frequencies of all neurons.

Brain states may be regarded as semantic symbols. The currently dominant view has it that these symbols can be decomposed, without loss of structure, into semantic atoms, and that these correspond to neurons (or small disjunctive sets of neurons). Each cell has attached to it a bit of symbolic meaning. The dynamically changing semantic symbol in the brain, according to this view, is additively composed out of the elementary symbols of all the active neurons. In the last two decades, neurophysiology has invested tremendous experimental efforts to detect the symbolic meaning of individually recorded cells and to discover regularities in their spatial distribution. Typical are sensory cells which respond to stimuli of a certain quality in a specific location of a sensory surface, or motor cells which command single muscles or muscle groups. More controversial cell types encode higher-level symbols, referring, for instance, to the visual image of a face or a hand.

The views sketched above, although they are the subject of continual discussion, nevertheless form the basis of virtually all of present-day concepts of brain function. However, neurobiology has been unable so far to account for even the most routine abilities of our brain within this conceptual context. In an earlier report [1] I have criticised the semantic symbol system sketched above as being inadequate, and have proposed a different interpretation of the neurobiological data, in terms of semantic symbols with a richer structure. The argument is summarized here and is complemented with some observations on recent progress with the dynamical problems involved. A more detailed account is forthcoming [2].

The semantic symbols of current view are fully characterized as being (non-ordered) subsets of the set of all cells in the brain. They thus are totally lacking hierarchical and syntactical structure. Such symbols are ambiguous in the sense that there is a virtually infinite number of ways in which they can be decomposed into combinations of subsymbols (see Fig. 1). The usual way of encoding subsymbol structure with the help of spatial arrangement of elementary symbols (e.g. of letters in the text) is not available in the brain, because its dynamics is controlled by the structure of the network and not the position of the cells (which is fixed anyway). Brain theories usually avoid the problem by a reduction or complete elimination of symbol decomposition ambiguity with the help of an appropriately specialized network structure. This subterfuge has, however, the disastrous consequence of eliminating all the functional flexibility which could flow from combinatorial construction of new high-level symbols from existing low-level symbols, together with the corresponding potential for economy and especially for generalization.
Superposition of symbols according to the conventional view. The box symbolizes some part of the brain, and is imagined to be filled with neurons. Each dot represents an active cell. On the left, there are two different symbols (referring, for instance, to two objects, seen at different times). On the right, the two symbols are superposed by simultaneous activation of both sets of cells (both objects are seen together). The rest of the brain can only react to the superposition, not the original subsets or subsymbols. False conjunctions of features will therefore lead to erroneous reactions.

**Structured Symbols**

In [1] it was argued that syntactical structure is introduced to the dynamical symbols in the brain with the help of additional variables to represent the binding or grouping of elementary symbols into subsets. On the fundamental level this may be done with the help of temporal correlations of nervous signals (to be evaluated with a resolution of a few milliseconds, and to be integrated over a few hundred milliseconds and longer). The necessary stochastic temporal structure in nervous signals is commonly observed, and also dynamically changing correlations have been reported [3, 4]. Correlations (in the sense of temporal synchrony) are known to have a strong influence on nervous dynamics, since neurons essentially are coincidence detectors: correlated signals converging on a cell have a much stronger effect than non-correlated signals (see Ref. [5] for a recent discussion). Subsymbols expressing their linkage by being correlated therefore can cooperate with their signals in influencing third parties, whereas non-linked subsymbols can not. Correlations can be processed by nervous networks as signals in their own right (cf. Ref.[6]). They can be induced in the network from an external source and they can be propagated. Correlations are spontaneously created in sets of cells which are synaptically coupled.

In order to dynamically control and stabilize correlations, synaptic links have to be modifiable on the fast time scale of fractions of a second [1] (see also Ref. [7]). This “synaptic modulation” is not to be confused with synaptic plasticity, which acts on a much slower time-scale. By synaptic modulation, the strength of an excitatory synapse, i.e. the size of the effect of an incoming pulse on the target cell, is controlled by the signals in the pair of neurons connected by
the synapse. Without nervous signals, synaptic strength slowly relaxes (within minutes) to a resting level, which is a certain fraction of the maximal strength of the synapse. Positive correlations (synchrony) increases the strength up to the maximum value, negative correlation decreases it, ultimately to the value zero. As to a discussion of a microscopic mechanism for synaptic modulation see Ref.[8]. There is a mechanism which suppresses global correlations and the activation of the full set of synapses in a network. This mechanism could be an inhibitory system which restricts activity in the network to a small fraction of cells at any one time. Synaptic modulation as controlled by signal correlations is a basis for rapid pattern formation in the brain. The necessary instability is the positive feed-back loop between correlations and synaptic strenghts. Different synaptic links cooperate with each other. For instance, a direct excitatory synapse between two cells cooperates with indirect excitatory links between the same cells. However, if two pathways from one cell to another differ too much in the number of intermediary cells (and correspondingly in temporal delay) the pathways cease to cooperate. Non-cooperating synapses are caused to compete by the suppression mechanism introduced above.

In order to fix ideas, let us introduce a simplified set of differential equations for the dynamics of synaptic modulation. It is assumed that there are \( N \) cells which all have the same average activity. The signals themselves are eliminated, and the only variables left are the excitatory synaptic strengths \( W_{ij}, \; i, j = 1, \ldots, N \), which are positive quantities. The main effects of signal correlations are taken into account implicitly, by noting that the pairs of synapses \( W_{ik}, W_{kj}, k = 1, \ldots, N \), cooperate with the synapse \( W_{ij} \). This is represented by a growth term with a kinetic coefficient \( \sum_k W_{ik} W_{kj} \). This sum happens to be an element of the square of the matrix \( W \). Global competition is introduced in the usual way [9–11] with the help of a growth limitation term involving a sum over all competing growth rates. Rows and columns of \( W \) are limited separately. The resulting differential equations are:

\[
\dot{W}_{ij} = W_{ij} \cdot (W^2)_{ij} \quad (\ast)
\]

\[-W_{ij} \cdot \left( \sum_{\not{j}} W_{i\not{j}} \cdot (W^2)_{\not{j}j} + \sum_{\not{j}} W_{ij'} \cdot (W^2)_{ij'} \right).\]

Terms with higher powers in \( W \) could correspond to cooperation of direct links with longer indirect links. They are left out for simplicity.

**Block Structures**

What is the structure of the connectivity patterns preferred by a dynamics of the type (\( \ast \))? The two terms on the right-hand side of equation (\( \ast \)) have a tendency to establish equality between \( W \) and \( W^2 \), and to normalize sums of rows and columns to 1. It is interesting to know that a positive symmetric matrix with these two properties has block structure (E. Bienenstock, personal
communication): under appropriate permutation of rows and columns the matrix has positive values in square blocks along the diagonal and is zero elsewhere, all values within one block being equal.

Let us assume for a moment that matrices with block structure were the preferred states of network dynamics. Would such dynamics be useful? The number of possible distinguishable states for a given set of \( N \) active cells would be rather large, even with given block structure. Let \( m_r, r = 1, \ldots, m \), be the size of blocks, \( \sum m_r = N \). The number of connectivity states is then \( M = N!/\prod m_r \). All these states would correspond to just one state in the conventional system. A typical dynamic process would start in an initial state in which a number of cells are active and communicate by a matrix \( W \). Most of the elements of \( W \) vanish because the corresponding synapses do not exist. The existing synapses may be in their resting state or they may already have been modulated in strength (e.g. by externally induced correlations). This initial structure of \( W \) (and, of course, also of the activity distribution) will now evolve dynamically, until a stationary state in the form of a decomposition of the set of cells (i.e. of the matrix) into blocks is reached.

It is important that under realistic conditions the blocks formed correspond to a useful segmentation into subsymbols. To demonstrate this, sensory segmentation has been treated as an application (C.v.d. Malsburg and W. Schneider, Ref. [17]). Local stimuli on a sensory surface (e.g. retina, cochlea) have to be grouped according to similarity in sensory qualities (e.g. colour, velocity, texture in the visual modality, inter-aural delay, amplitude modulation, frequency modulation, harmonic structure in the auditory modality). In this way the system can find segments which correspond to coherent objects in the external world. With the help of quality-specific cells and an appropriate array of synaptic connections it can be achieved that in the initial state produced by an external stimulation, \( W \) already has a weak block structure. Feedback between the ensuing signal correlations and synaptic modulation soon leads to a clear-cut block structure. As a consequence, cells responding to parts of different objects are temporally anticorrelated and thus are prevented from interfering with each other during pattern recognition. The simulations referred to above were done with a formulation of dynamics somewhat more complicated than (†), treating signals and modulating synaptic strengths as separate variables.

The blocks formed in a first stage of organization may be unstable and may decompose into smaller blocks. This can go on through a number of stages until blocks of a certain minimal size are formed. It is not clear yet how this type of instability can be handled. There are several ways to stabilize blocks, but some of them may unduly limit the range of symbols which can be formed. If the system is regulated such that connections between blocks are only weakened and bonds are not ruptured completely, a hierarchical system of blocks and sub-blocks may be formed. (The corresponding matrix structure would resemble the one invoked in the replica symmetry breaking scheme of spin glass theory, Ref. [12].) One may conjecture that such a regime can be found by appropriate
formulation of dynamic interaction terms and by appropriate choice of control parameters, and that the actual dynamics of the brain is of this type. Such a system is ideally suited to form hierarchically structured semantic symbols.

**Topological Connection Patterns**

Given a set \( S \) of \( n \) cells, a metric space \( T \) of low dimension (say 2 or 3), and a mapping \( f \) from \( S \) to \( T \), a topological connection pattern is obtained by connecting each of the \( n \) image points in \( T \) with a small number of closest neighbours. I have previously [1] formulated the conjecture that under appropriate control topological connection patterns are preferred states of network dynamics.

There are several lines of argument to justify this conjecture. Topological graphs are sparse and have many short loops (the term “graph” refers to an abstract connection pattern, i.e., without reference to cellular identity). They thus conform to the two conflicting tendencies inherent in the dynamics as formulated, for instance, in \((*)\). Any graph can be interpreted as a topological graph in a space with sufficient dimensionality. The above conjecture therefore is equivalent to the statement that network dynamics reduces the dimensionality needed to harbour the topological graph. The dimensionality of a graph can be defined as the power with which the total number of descendants of a node grows asymptotically with generation number (if there is such an asymptotic power law). Connectivity dynamics prefers situations in which different paths emanating from a node cooperate with each other by converging on other nodes. With other words, dynamics prefers situations with a minimal number of descendants. This leads to a reduction in the dimensionality of the graph.

Topological graphs may be constructed from block structures in the following way. Let \( f \) be again a mapping of set \( S \) of \( n \) cells into the metric space \( T \). Let there be a family of overlapping tesselations of \( T \) into smaller regions. Each tesselation divides the image points of \( S \) into a number of blocks, for which there is a corresponding matrix which contains connections only within the blocks. Superpose all matrices belonging to the family of tesselations and normalize the resulting matrix appropriately. The result corresponds to a topological graph.
A given topological graph can be realized in a number of ways as connection pattern of a given set of cells. Two different permutations $P$ and $P'$, applied to the set $S$ after it is mapped to $T$, in general yield different connection patterns. This is illustrated in Fig. 2. Several such patterns can be superposed on the same set of cells. In general, a superposition of topological connection patterns is not itself topological in structure. The network therefore has to respond dynamically, e.g. by the temporary amplification of one of the patterns in the superposition and by the deactivation of all others. This process has been demonstrated in simulations both of a system of autonomous $W$-equations similar to $(*)$ (together with K. Schulten) and of a neurophysiologically more realistic formulation (A. Häussler and C.v.d. Malsburg, Ref. [18]). In the latter, cells are dynamically described by a membrane potential which evolves due to

**Fig. 2**

Superposition of two topological connection patterns. $S$ is a set of 12 cells, numbered from 1 to 12. In a they are mapped in correct order into a circle in the two-dimensional plane. Each cell is connected to its four nearest neighbours. In b the cells have been permuted after application of the same mapping and before they are connected to neighbours. In c the two connection patterns of a and b have been superposed. Realistic patterns connect huge numbers of cells by many more (unidirectional) connections per cell, and may be less regular.
linear decay, to noise, and to synaptic currents obtained in the usual way as the sum of incoming signals, weighted with the actual values of synaptic weights. The membrane potential saturates at the Nernst potentials of the ion species which compose the excitatory and inhibitory currents. The inhibitory system is modeled by a single pool which monitors and limits the total activity of all excitatory cells. A cell produces an action potential if its membrane potential exceeds a threshold. After a pulse, the cells are made refractory by a temporary increase in threshold. In the initial state, connectivity corresponds to the superposition of two topological patterns in the form of cyclic graphs which connect neighbors and next nearest neighbors, see Fig. 3. Fig. 4 shows a sample of activity with this initial connectivity. Each component pattern tries to impose its preferred type of activity, circularly running waves. The two connection patterns, being related by a permutation, are in conflict with each other. Synapses are modulated according to a control function which is positive when there is a coincidence of activity in the presynaptic and postsynaptic cell, and which is negative when only one of the signals is active. (Small delays are tolerated for synchronous signals.) After some synaptic reorganization, the matrix has the form shown in Fig. 5, and the signals, see Fig. 6, now have the form of a simple running wave. The simulation shows that indeed a non-topological graph of the type shown in Fig. 2c is dynamically reduced to a topological one, like the one in Fig. 2a. This type of “network reduction” may be one of the most important dynamical processes in the brain.
Connectivity matrix of the superposition of two topological connection patterns. The entry in row \(i\) and column \(j\) represents the strength of the synapse from cell \(i\) to cell \(j\). Both patterns have a structure analogous to the one in Fig. 2a. A simultaneous permutation of rows and columns would bring the scattered weights to fall on a diagonal band (the original band being scattered over the matrix). One of the patterns has 10\% advantage over the other (strength 70 as against 63).
Activity in the network of Fig. 3, with the dynamics described in the text. Columns correspond to cells. Time runs downward, each row of dots representing a moment in time. Nervous impulses are indicated by vertical strokes. Each connection pattern tries to impose its preferred activity pattern, running waves. The two patterns are, however, in conflict with each other.
Fig. 5
Connectivity matrix after reorganization by synaptic modulation. One of the ring-shaped patterns has won the competition and has increased in strength, to values near 1.0. The other pattern has been deactivated, with the exception of some synapses "near" the winning pattern.
preferred activity pattern, circularly running waves. If, on the other hand, another ring-structure (as in Fig.2b) is activated on corresponding points. For the ring structure of Fig.2a, this is shown in Fig.7a. If one of the connected rings is activated, it is easy for it to activate also the other. If, on the other hand, another ring-structure (as in Fig.2b) is active on the left set of cells, the same projection of fibres cannot activate the right-hand ring, see Fig.7b (especially in a competitive situation), and the projection will soon be deactivated. This example shows that the structure of internal cor-

Application of Topological Connection Patterns

Topological patterns can be connected such that the result has again topological structure and is preferred by the dynamics. From two homeomorphic topological patterns, for instance, a new one can be formed by connecting corresponding points. For the ring structure of Fig.2a, this is shown in Fig.7a. If one of the connected rings is activated, it is easy for it to activate also the other. If, on the other hand, another ring-structure (as in Fig.2b) is active on the left set of cells, the same projection of fibres cannot activate the right-hand ring, see Fig.7b (especially in a competitive situation), and the projection will soon be deactivated. This example shows that the structure of internal cor-

Fig. 6
Activity in the network of Fig.5. The winning connection pattern now is free to impose its preferred activity pattern, circularly running waves.
relations and active connections of a set of cells is important to control the projection of activity in this set to other sets in the brain. Another way of joining two $d$-dimensional topological patterns is by sewing them together along $(d-1)$-dimensional borders.

If cells are considered as semantic atoms, connection patterns are to be interpreted as syntactical structures. In the literature, a wide variety of syntactical structures is discussed. Artificial intelligence and linguistics often use hierarchical list structures and graphs in the form of trees. The block structures and topological graphs discussed here have an enormous advantage of being able to serve as a basis for self-organization. It is not necessary to have specific algorithms in order to process specific syntactic structures, as in the artificial intelligence approach. There is only one universal “algorithm”, network dynamics, which ensures that each initial state converges towards the structure which is most consistent in itself and with boundary conditions. As to the usefulness of these universal syntactical structures in practical situations there cannot be a simple argument, to be sure. It is, however, possible to demonstrate this usefulness with the help of examples, of which some are now in the process of being worked out in detail. Two examples have already been mentioned above: segmentation of a sensory pattern into blocks of structurally related elements, and projection between structures in the brain on the basis of homeomorphy of syntactical structure. A third example is discussed now.
Projection between topological connection patterns. The oval-shaped rings symbolize patterns analogous to the one of Fig. 2a. Only two connections per cell within the pattern are shown for clarity. In each case two homeomorphic patterns are connected by a one-to-one projection. In 7a the projection corresponds to a homeomorphic mapping: pairs of connected cells project to pairs of connected cells. The resulting “superpattern” is of the topological kind again, and is stabilized by network dynamics. In 7b another topological connection pattern is active on the left set of cells. The fibre projection now does not correspond to a homeomorphism: pairs of connected cells now project to pairs of non-connected cells, in general. Activity patterns shaped by the left net do not resonate with the ones in the right net, and activity is not easily transferred. In addition, the projection will soon be deactivated by synaptic modulation.
How to disambiguate Rosenblatt’s model for invariant pattern recognition. Cells in layer $A^{(2)}$ respond to features irrespective of their position on the sensory surface $S$. For each feature-cell $\alpha$ there is a class of cells $(\alpha,x)$ in $A^{(1)}$, one cell for each position $x$. Cell $\alpha$ receives connections from all cells $(\alpha,x)$, and can be fired by one of them. In this way, an activity pattern in $S$ activates a set of feature cells in $A^{(2)}$. This set, by construction, is invariant with respect to the position of the sensory pattern. However, it is ambiguous. The ambiguity can be reduced if cells in $A^{(1)}$ encode their neighbourhood relationships by signal correlations (produced by common retinal origin or by short-range connections.) The correlations are propagated to $A^{(2)}$, where they activate a connection pattern (dashed lines) which encodes the topology of $A^{(1)}$. 

Fig. 8
Invariant Pattern Recognition

Our brain routinely associates the visual image of objects to information from other modalities (e.g. motor command, language). One and the same object may be represented on the retina by an infinite variety of patterns of different size, position, orientation, and so on. In order to be able to generalize from one such pattern to the others it is necessary for the brain to organize an invariant representation for an object. Invariant pattern recognition is an outstanding unsolved problem both for science and for engineering. (In general, a scene is composed of a number of different patterns. Before any one of them can be recognized, the scene has to be decomposed into segments which are likely candidates for recognizable stored patterns. It is assumed here that the problem of segmentation has already been solved, presumably in the way indicated above.)

There are two extreme attempts at the problem. One of them uses template matching, the other is based on extraction of invariant features. A template is a canonical instance of a pattern. The pattern to be recognized is transformed as a whole until it matches one of the stored templates. Template matching is impractical in realistic situations because the space of transformations to be searched is too large, and because it is often not clear in advance which segments of a scene are to be considered as recognizable patterns. The invariant feature approach is best explained with reference to Rosenblatt’s four-layer perceptron [13], see Fig.8. There are three “layers” of cells (the fourth layer has no relevance here), \( S, A^{(1)}, A^{(2)} \). \( S \) corresponds to the retina. Each cell in \( A^{(2)} \) responds to the occurrence of a small subpattern (“feature”) of a type specific to the cell, without regard to where it appears on the retina. This is made possible with the help of extensive machinery in layer \( A^{(1)} \). To a feature of type \( \alpha \) in \( A^{(2)} \) there corresponds a whole class of cells in layer \( A^{(1)} \), each of which is specific to the feature \( \alpha \) in a fixed position. Each \( \alpha \)-cell connects to, and can trigger, the one \( \alpha \)-cell in \( A^{(2)} \). When a specific pattern is presented to \( S \), the subset of all those cells in \( A^{(2)} \) is activated the features of which are part of the pattern. By construction, the response in \( A^{(2)} \) is invariant with respect to translation. (The scheme can easily be generalized to invariance with respect to wider transformation groups.) Unfortunately this response is ambiguous in general, because many patterns of different structure may be composed of the same set of local features, arranged differently. Therefore, also this second approach to invariant pattern recognition is impractical. The weakness of the invariant feature approach stems from the fact that all information regarding spatial relationships between features is discarded on the way from \( A^{(1)} \) to \( A^{(2)} \) along with position information. It has been proposed earlier [1] that the syntactical structure of neighbourhood relationships within \( A^{(1)} \) could be encoded as temporal correlations of cellular signals, and that they could be transferred in this form to \( A^{(2)} \). The connections within \( A^{(2)} \) would then be modulated by these correlations, and soon the pattern of synapses activated in \( A^{(2)} \) would express neighbourhood relationships within \( A^{(1)} \) and within \( S \).
In this way the ambiguity of the activity pattern in $A^{(2)}$ could be reduced or eliminated. The structure of the net in $A^{(2)}$ would control projections to other structures in the brain, in the way indicated above. The set of neighbourhood relationships in $A^{(1)}$ evidently has the topological structure of the 2-dimensional plane. Correspondingly, also the net activated in $A^{(2)}$ would have this same structure, although for a given set of cells in $A^{(2)}$ different patterns activate different topological nets (related by permutation). Accordingly, the whole net to be formed in the ensemble of active cells in $A^{(1)}$ and $A^{(2)}$ is of the topological type and hence corresponds to a dynamically preferred structure.

**Conclusion**

After the great optimism of the cybernetical era one or two decades ago, brain theory has had a period of stagnation. In response, important conceptual adaptations have taken place, characterized mainly by a transition from an algorithmic to a dynamical description of neural systems. In the algorithmic scheme, a process is controlled deterministically by a program (or a rigid network) which is preexistent: one may speak of hetero-organization. Modern dynamical formulations put emphasis on probabilistic mechanisms and random search, in order to implement auto-organization [14, 15, 16]. In this development, the analogy to physical systems (especially thermodynamical systems) has played a prominent role, and the theoretical tools employed by physicists to describe dynamical systems have become important.

It was the purpose of this paper to report on progress with a further conceptual adaptation, concerning the functional interpretation of brain states, or the nature of the dynamical object. It is argued that this dynamical object is not just a vector of activity levels in a given set of units, but comprises in addition a dynamical graph of connections between those units. A dynamical graph, in addition to being a natural basis for organization, is a prerequisite for the implementation of a flexible symbol system. Definite proof of the validity of this approach can only come from large-scale simulations, and especially from demonstration of such important processes as pattern recognition, scene analysis and sensory-motor coordination in truly parallel machines constructed on the basis of principles advanced here.

This work has in part been supported by a grant of the Stiftung Volkswagenwerk.

**References**


Presented at the Discussion Meeting of the Deutsche Bunsen-Gesellschaft für Physikalische Chemie “Dynamically Organized Systems” at Schloß Elmau, October 15th-17th, 1984

©VCH Verlaggesellschaft mbH, D-6940 Weinheim, 1985 0005–9021/85/0606–0703 $ 02.50/0

18