Self-Organization and the Brain

Christoph von der Malsburg

Introduction

The process of network self-organization is fundamental to the organization of the brain. It takes place on several temporal scales: the ontogenetic/learning time scale of hours, days, and years; and probably also the functional time scale of fractions of a second to minutes (see Dynamic Link Architecture). The basic concepts of network self-organization are discussed here. More extended introductions to the topic of self-organization are found in Prigogine and Stengers (1984) and Murray (1993), for instance.

One often speaks of some structural trait of an organism as being "genetically determined." This seems to imply that the genes contain a blueprint describing the organism in full detail. However, all the stages of brain organization (not just evolution) more or less strongly involve an element of self-organization and creativity. It has often been emphasized that the genes cannot, in any naive sense, contain the full information necessary to describe the brain. The cerebral cortex alone contains at least on the order of 10^{10} synapses. Forgetting considerations of genome size, one can hardly imagine how ontogeny could select the correct wiring diagram out of all of the alternatives if

this strategy has the possible disadvantage of requiring a serial analysis of a scene. It results in a tremendous savings in computational resources.

Road Map: Vision
Related Reading: Dissociations Between Visual Processing Modes; Figure-Ground Separation; Thalamus

References

all were equally likely. Besides, judging from the variability of the vertebrate brain structure, the precision of the ontogenetic process is not sufficient to specify individual connections.

The conclusion one must draw is that ontogeny makes use of self-organization, that is, of general rules to generate neural structure and of principles of error correction. Above all, ontogeny can only produce structures with a high degree of regularity—for example, homogeneity, repetitivity, or continuity. Knowing the mechanism of ontogeny is of extreme importance: one cannot understand the function of the brain without knowing its structure, and one cannot know the structure of the brain without knowing the principles of its ontogenesis.

Abstract Scheme of Organization

There are well-studied paradigms of pattern formation, especially in physics, physical chemistry, and astronomy: convection, crystallization (or more generally, phase transitions), reaction-diffusion systems (the emergence of spatial and temporal chemical patterns, e.g., in the Zhabotinski-Belousov reaction), and star and galaxy formation. I will attempt to give here a general description of the basic mechanisms of organization by using the important example of convective pattern formation, the so-called Bénard problem (see Prigogine and Stengers, 1984).

Organization takes place in systems consisting of a large number of interacting elements. These could be atoms in a liquid or crystal, or small subvolumes of liquid in convection currents. In a reaction-diffusion system or an evolving star system—or, in the application that is of interest here, these would be synapses in nerve networks. Initially, self-organizing systems are in a relatively undifferentiated state: atoms move randomly and all subvolumes of the liquid are in the same state of motion or have the same chemical composition. Then, some small, typically random deviations from that state arise; for example, some convective fluid motion sets in. To stress the random nature of typical small deviations, they are called fluctuations.

In the prime example, the Bénard phenomenon, a flat vessel is filled with liquid and its bottom is homogeneously heated. As long as the temperature gradient is below a certain threshold, heat is conducted from the lower to the upper surface without bulk movement of the liquid. However, above that threshold, the warmer, lighter liquid near the bottom rises and cooler liquid from the top flows down. Under homogeneous conditions, this flow pattern is very regular and has the form of hexagons or rolls.

From this and many other organizing systems, the following three principles may be abstracted:

1. Fluctuations self-amplify. This self-amplification is analogous to (asexual) reproduction in Darwinian evolution. In the Bénard system, fluctuations are created by thermal motion. If a small column of liquid moves upward, more warm liquid is drawn in from the bottom, the column becomes less dense, and its upward movement is accelerated. Downward movement accelerates analogously.

2. Limitation of resources leads to competition among fluctuations and to the selection of the most vigorously growing (the "fittest") at the expense of the others. In the Bénard system, upward movement in one place requires downward movement in other places. The columns with the least density will win and rise.

3. Fluctuations cooperate. The presence of a fluctuation can enhance the fitness of some of the others, in spite of the overall competition in the field. (In many systems, the "fitness" of a fluctuation is identical with the degree of cooperation with other fluctuations.) The liquid near a column of rising liquid is dragged up by viscosity.

The identification of these three principles with features of a concrete system is sometimes ambiguous. In the Bénard system, competition in terms of upward movement might also be seen as cooperation between upward movement occurring in one place and downward movement occurring in another place. Whole coherent patterns of movement, again, compete as long as there is local contradiction between them: liquid cannot move up and down at the same place.

A fundamental and very important observation about organizing systems is the fact that global order can arise from local interactions. Many originally random local fluctuations can coalesce into a globally ordered pattern of deviations from the original state. The intermolecular forces acting within a volume of liquid are of extremely short range, yet the patterns of convective movement they give rise to may be coherent and ordered on a large scale. This fact will be one of extreme importance to the brain, in which local interactions between neighboring cellular elements create states of global order, ultimately leading to coherent behavior.

The stage for the organization of a pattern is set by the forces between elements and by initial and boundary conditions. In the Bénard system, these forces are the hydrodynamic interactions, gravity, thermal conduction, and expansion. Boundary conditions are set by temperatures at the upper and lower boundary and by the form of the vessel. In the nervous system, the stage for the generation of connection patterns is ultimately set by prespecified rules for the interaction of cellular processes and signals, and by the environment. Because nerve cells are connected by long axons, there is an important and exciting difference between the nervous system and most other examples studied so far. Neural interactions are not necessarily topologically arranged; connected cells are "neighbors" although they may be located at different ends of the brain. This gives rise to genuinely new phenomena. Some of the ordered structures within the nervous system may not "look" ordered to our eye, which relies essentially on spatial continuity. However, in the concrete cases considered here, ordinary space will still play a dominant role.

An organizing system may contain a symmetry such that there are several equivalent organized patterns. These compete with each other during organization. In the Bénard system, if set up in a circular pan, any organized pattern could be rotated around the center of the pan by an arbitrary angle to obtain another valid pattern. One of these has to be spontaneously selected during pattern formation, a process that is called spontaneous symmetry breaking. When the boundary or initial conditions are slightly deformed, so that the original symmetry is destroyed, one organized pattern is favored. In general, self-organizing systems react very sensitively to symmetry-breaking influences.

Neural Network Organization

Two types of variables are relevant to network organization: signals and interconnections. Signals are the action potentials that are propagated down the axonal trees of neurons. Connections control neural interactions and are characterized by weight variables. These measure the size of the effect exerted on the postsynaptic membrane by arriving nervous impulses. Correspondingly, organization takes place on two levels: activity and connectivity.
On the ontogenetic time scale, one is interested mainly in network self-organization, which has the following general form. Assume that previous processes have already set up a primitive network. This network, together with input signals, creates activity patterns, and these activity patterns in turn modify connections by synaptic plasticity. The feedback loop between changes in synaptic strengths and changes in activity patterns must be positive, so that coherent deviations from the undifferentiated state self-amplify, conforming to the first of the principles previously formulated. The process is constrained by the requirement that modifications in a synaptic connection have to be based on locally available signals. These are the presynaptic signals, the postsynaptic signal, and possibly modulatory signals that are broadcast by central structures. The postsynaptic signal could be a local dendritic signal or the outgoing axonal signal.

The requirements of self-reinforcement and locality suffice to specify the mechanism of synaptic plasticity in excitationary synapses: A strong synapse leads to coincidences of pre- and postsynaptic signals which, in turn, increase the strength of the synapse. Hebb (1949) gave this formulation:

When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

This rule is referred to as "Hebbian plasticity" (see Hebbian Synaptic Plasticity). The corresponding rule for inhibitory synapses would have a synapse strengthened if it was successful in inhibiting the postsynaptic element. At present, however, most authors consider inhibition as a rigid service system that does not take part in network self-organization.

Hebb's rule corresponds to the "self-reproduction" of the general scheme of organization. To stabilize the system, some competition for limited "resources" has to be introduced. Most likely, there is a mechanism of isostasy by which each cell keeps the temporal average of its activity (taken over the span of some hours) constant. As a consequence, the increase in strength in some synapses must be compensated for by a decrease in others. Only the more successful synapses can grow; the less successful ones weaken and eventually disappear. For technical reasons, some models discuss a simpler competition rule for synapses, in which the sum of the synaptic weights of all synapses converging on a cell is kept constant. This rule leads to certain functional deficits and is probably not realistic. Synaptic plasticity, constrained by competition, implements organizing principles 1 and 2.

One synapse on its own cannot efficiently produce favorable events. For that it needs the cooperation of other synapses that converge onto the same postsynaptic neuron and that carry coincident signals. This implements the third organizing principle. In order for such coincidences to occur consistently, there must be a causal connection between presynaptic cells. Synaptic plasticity is the means by which the nervous system detects such causal connections. Coincidences may result from excitatory links between presynaptic neurons. They may, however, also be caused by simultaneous stimulation of sensory cells, in which case they point to the existence of causal connections in the external world.

The rules of cooperation and competition act on a local scale. The phenomenon of self-organization is the emergence of globally ordered states, as discussed in context with the emergence of global convection patterns in the Benard phenomenon. The term global order is used for configurations that bring the local rules into a state of optimal mutual consistency with each other. The fact that the external world takes part in the game leads to the adaptation of the nervous system to it.

The rules for the adjustment of synaptic weights that have been introduced are able to produce ordered connection patterns. However, they do not necessarily organize the nervous system for optimal biological utility. For this, two types of controls are necessary: (1) genetic control of boundary conditions and interaction rules to favor certain useful connection patterns; and (2) control by central structures that are able to evaluate the degree of biological desirability of activity states. If a state proves to be useful, a gating signal is sent to all of the brain, or to an appropriate part of it, to authorize synaptic plasticity. That state is thereby stabilized, and the likelihood for its future appearance is increased.

Central control as the only criterion for growth or decay of synapses is not sufficient. Assume our nervous system evaluates the usefulness of its state once per second. It then could create less than $3 \times 10^9$ bits of information in our lifetime, for that is about the maximum number of seconds given to us. This certainly is not sufficient to regulate the strengths of all of the $10^{12}$ synapses of our cerebral cortex. On the other hand, this amount of information may be sufficient to select from among the relatively small universe of ordered connectivity patterns that can be created by rules of local cooperation and competition under predetermined constraints.

Conclusion

The last two or three decades have seen a revolution in thinking about organization and the origin of structures. This revolution has swept across all fields of human thinking and is also deeply affecting our view of brain and mind. In the prerevolutionary view, organization is the result of a preexisting plan, which has arisen outside the field of study in some separate agent. This plan is realized with the help of some explicit mechanism, and the study of this mechanism is the study of the process of organization. You may refer to this scheme by the word hetero-organization. In the post-revolutionary view, there is no preexisting plan. But there is a dynamical system of interacting elements which spontaneously fall into globally ordered patterns. Thus, the "plan" can only be read off the final product of the process of organization and is in general not accessible ahead of time.

There is a number of mathematical tools to study self-organization in a systematic way. Among them are systems of nonlinear differential equations and their analysis with the help of analytical methods, stability analysis and bifurcation theory, and numerical simulation. Some of these methods are introduced in Murray (1993; see also Pattern Formation, Biological). An altogether different approach is represented by the methods of statistical mechanics and phase transition physics: for a general introduction, see Prigogine and Stengers (1984).

The general principles of self-organization in neural networks and other systems presented here are applied and illustrated in various articles of this Handbook. Prime examples are the formation of cortical domains (stripes or blobs or barrels) and the establishment of retinotopic mappings or of orderly arranged orientation sensitivity in the visual cortex (see Development and Regeneration of Eye-Brain Maps and Ocular Dominance and Orientation Columns). Great challenges still lie ahead, especially a more detailed understanding of the ontogenesis of the brain and nervous system, the ontogenetic construction of areas and their connective architecture, the construction of the functioning brain by growth and learning.
Self-Organization in the Time Domain

John G. Taylor

Introduction

Besides space, time is a crucial aspect of incoming information. There are many situations in the animal world where the ability to process, store, recognize, or recall temporal sequences of patterns has great survival value to the animal. At the human level, language has allowed for the efficient handling of high-level concepts. In this article we will not consider hardwired delay-line systems, but restrict ourselves to adaptive neural networks able to learn the temporal features on inputs.

In order to achieve learning of such temporal structures, there may be an ability, in either the net itself or in an auxiliary structure, for the temporal character of the input to be buffered temporarily. This buffering might be achieved directly by reasonably long time constant neurons on a "history" net, as in the net L of Figure 1.

Leaky integrator neurons (LINs) have activity on their surface at a given time which dies away with a certain time constant. This decay of past activity corresponds to the effect of new activity being reduced by a factor (less than one) at each subsequent time step. The decay time constant is related to this reduction factor in a simple manner. Experimental data on cortical neurons show that biologically realistic values of this decay time for a simple neuron will have a maximum of about 50 milliseconds. However, the presence of integrating circuits with a time constant of up to 20 seconds for the vestibuloocular reflex indicate that model neurons with a similarly large time constant may be taken as a first approximation to more complex underlying microcircuitry.

It is important that these model LINs have a range of time constants, so that a history of the past inputs is almost in a one-to-one correspondence with the activity stored in the buffering net L of Figure 1. This activity is then stored and used to guide output of the further network M of Figure 1 to produce the next input. This leads to the method of predictive self-learning, which will be discussed in more detail after the more complete description of temporal neurons in the next section. Alternatively, the changing nature of the patterns can be learned directly by comparison of new with ongoing net activity. The adaptation can be done using Hebbian learning in the case, say, of a recurrent Hopfield net (Amit, 1989) to learn the transition between sequences of patterns. This might be called the temporal Hopfield map. One may also use a Mexican-hat style of lateral inhibition to produce a topographic map, but with temporal neurons holding a history of past activity. This method is appropriately termed the temporal Kohonen map. This approach can also be improved by storage of histories of inputs on separate synapses leading to traces and improved sequence storage (Critchley, 1994). The method and these extra features will be discussed in the section "Temporal Topographic Maps."

There have been numerous other approaches to temporal sequence storage and generation using temporal neurons, for example, the outstar avalanche approach of Grossberg. This does not use a range of time constants for the neurons, so may have problems with the disambiguation of sequences which have repetitions in them (such as the sequence ABCCABD). There is also the fact that loss of the feedback structure of hippocampus causes amnesia for episodic memory (see SHORT-TERM MEMORY); neurobiological realism therefore supports the use of feedback architectures for some aspects of temporal sequence learning, rather than simple feedforward systems.

Temporal Neurons

Leaky integrator neurons act by storing activity coming onto their surface, this activity \( A(t) \) at time \( t \) dying away with certain time constant. Thus in discrete time

\[
A(t+1) = (1-d)A(t) + I(t)
\]

where \( (1-d) \) is the reduction factor experienced at each time step, and \( I(t) \) is the input at that time. Further temporality could be obtained by including nontrivial geometry for the