Atti del Congresso *Nuovi problemi della logica e della filosofia della scienza*. Viareggio, 8-13 gennaio 1990.
Vol. I. CLUEB, Bologna (Italy) © 1991.

LOGICS AT DIFFERENT LEVELS IN THE BRAIN

Valentino Braitenberg

Max-Planck-Institut für biologische Kybernetik
Spemannstrasse 38, D-7400 Tübingen, FRG

1. What is logic from the point of view of the brain theorist?

The laws of thought: but this is what we are after anyway - perhaps even with more relevant knowledge.

Is it the mathematical form that is given to what apparently governs the thinking process? This would be welcome: we are groping for a language.

Or is it the foundation of mathematics, in which case we have little in common.

Superficially, it is algebra, with binary variables. This was the first spark of recognition between brain theory and logic: at several levels in the hierarchic organization of brain functions there are elements which can be effectively described as having only two states.

I don't even know whether in an abstract definition of logic the binary variables are essential. I suspect not.

2. Binary variables in the brain

Sherrington, about a century ago, examining the combined effects of the electrical stimulation of several afferent nerves on the electrical activity in an efferent nerve of the cat spinal cord (each of these nerves containing many thousand fibers) was the first to introduce something like logic into neurophysiology.

When the input to two nerves was combined, the output (measured in a way that roughly represented the number of neurons activated) was sometimes more than the sum of the output to the two nerves separately, sometimes less, and sometimes even less than the output after stimulation of only one of the two nerves. These different situations were called facilitation, occlusion, and inhibition respectively. The interpretation that was offered was in terms of neurons in the spinal cord receiving input from both nerves. When there was facilitation, one could suppose that some of these neurons responded only when they were excited by both: they would be added to the output in case of combined stimulation. Occlusion was interpreted as evidence for neurons reached by both nerves and sufficiently excited by either; with combined stimulation, they would appear only once in the sum and their number would be subtracted from the sum of the reactions to separate stimulation. Finally, inhibition was taken as evidence of one nerve preventing the activity of neurons otherwise excited by the other nerve.

The macroscopic observations dating back to the last century held up to later detailed investigations of single neurons with microelectrodes. There are indeed two opposite kinds of influence which neurons can exert onto each other, excitation and inhibition, and the effects of excitation are so catastrophically non-linear that it does make sense to talk in terms of a threshold of excitation beyond which the unitary signal, the "action potential" is triggered, while for excitation which does not meet the threshold, the effects are minor, short lasting and do not propagate.

3. McCulloch and Pitts theory

Thus neurons could be stylized as "threshold elements" and it was easy to show that the "firing" of a neuron (= its action potential) in a network corresponds to a proposition on the state of its input neurons being active. Chains of neurons could be imagined to give temporal depth to these propositions and reentrant chains provided the possibility of letting neurons correspond to propositions referring to an indefinite time in the past. This was McCulloch & Pitts' theory (1943), much quoted by philosophers and automata theorists (Shannon and McCarthy, 1956).

There are two reasons for which this theory is unlikely. The first is the problem of timing. For a network of McCulloch & Pitts type neurons to process information, strict temporal order must govern the passage of signals from one neuron to the next, all the times of activation being multiples of some unit of time. Otherwise the coincidence of signals at some junction, which is required for it to function as a logical gate is not guaranteed. The truth is that no such pace-maker is found in the brain. Worse, the time which elapses between the arrival of an afferent excitation and the activation of the neuron depends on the amount of excitation exceeding the threshold: the more, the shorter the delay. Contrary to what happens in the abstract threshold element, the information residing in the intensity of the excitation beyond threshold is not lost, but is translated in a temporal signal, and the activity of the network falls out of step.

The second reason is that in the many years in which the McCulloch-Pitts idea served as a guide to neurophysiologists, no neuron was found whose activity corresponded to any meaningful event or thing in the world. The suspicion is now that "things" or "events" correspond in the brain to more complex patterns of activity.

4. Carriers of meaning in the brain

The question of what is a thing or an event is not one of logic, I presume. It is related to the concept of *morpheme* in linguistics. As in language there is an alphabet of elements (phonemes or perhaps "distinctive features" of phonemes) below the level of meaning, the carriers of meaning (morphemes) resulting from their combination, in the brain, too, such a two-stage representation seems to be at work. If the activity of individual neurons does not correspond to morphemes, their combinations most likely do.

There is good evidence on the nature of the elementary signals which are relayed by individual neurons.

Immediately attached to sense organs we find neurons which respond directly to the reception of the kind of energy to which the sense organ is tuned. The response is mostly monotonic, although hardly ever simply proportional. Mostly there is saturation beyond a certain intensity of stimulation. Frequently the response reflects the intensity of the stimulus as well as its change in time. Sometimes the response is negative, the neurons being inhibited by the stimulus.

The intensity of the response is given by the frequency of the action potentials produced by the neurons, as well as by the duration of the burst in which they occur. In most cases the intervals between individual action potentials in the burst, although quite variable, do not seem to act as carriers of information.

One step removed from the sense organs we find neurons which code slightly more complex signals, characterized by certain spatial and temporal patterns of an elementary nature. Some respond to a white spot surrounded by black, or vice versa, or to a spot of colour surrounded by a different colour, to the coming in sight of such a spot or to its disappearance from sight. In the acoustic system it may be a sound of a certain frequency, or a noise, or the cessation of a noise.

At the level of the cerebral cortex, which is sometimes considered the "highest" in the nervous system, things are only slightly more complicated when individual neurons are observed. They may respond to movement in a certain direction of the visual field, or to a contour of a certain orientation, to an acoustic frequency modulated upward, or downward. There have been reports of neurons responding to the sight of a human hand, or of a face, but mostly the responses even in the cortex remain below the level which in language acquires the dignity of a morpheme.

Finally, near the motor output, neurons code or rather command the state of contraction of a muscle, or the coordinated activity of several muscles which move a joint in a certain direction: again entities at a level more elementary than the units of meaning in motor action.

In the central and in the motor neurons too, like in the sensory ones, it is always bursts of rather disarrayed action potentials, rather than single action potentials, that correlate with the stimulus or with the response, and the parameters of the burst indicate something like intensity or distinctivness of the stimulus or the response.

5. Cell assemblies

All of this makes McCulloch-Pitts theory in its original form look rather unrealistic today. And yet, there is much in language that looks like a logical network with binary variables (a morpheme or a phoneme is understood or not, is uttered or not, a rule of grammar applies or not) and there is much in cognition which can be stylized that way. Apparently we must look for discrete entities at a level above that of the activity of individual neurons.

Such entities had been postulated early (1949) by psychologists (Hebb) and were termed cell assemblies. These are groups of neurons which are held together by reciprocal excitatory influences. They may be partly overlapping and still remain distinct, as long as the activitation of a sufficient number of neurons pertaining to one cell assembly leads to the activation of that assembly in its entirity and not of the others. This property was particularly pleasing to perceptual psychologists who had long seen in the "completion of patterns" or "reintegration" as it was also called, one of the basic phenomena in perception: partial evidence lets us perceive the whole thing.

The positive feedback within a cell assembly, leading to its explosive ignition, may well be the material counterpart to the binary variables of logic, existence or not existence, truth or falsehood and the like.

The postulated cell assemblies were supposed to be assembled through experience. A special kind of synapses (= contacts between neurons) was postulated which make the influence of one neuron onto another the stronger the more often the two have been active together. Thus elementary properties, represented by neurons, are tied together, when they belong together, into the things and events of one's experience, represented within the brain by cell assemblies.

This was all mere speculation, until it was shown by histological analysis (Braitenberg, 1978a,b; Braitenberg & Schüz, 1989) that the structure of the cerebral cortex fits admirably the kind of network one would like to postulate as a substrate of cell assemblies. Most of the synapses there are excitatory, and are of a kind which is most likely "plastic", or "Hebbian" i.e. apt to change with experience. Some direct observations with microelectrodes on neurons whose inputs were driven in a correlated or non-correlated way (Wiesel & Hubel 1965, Hubel & Wiesel 1965) showed that indeed correlation is translated in the cortex into strength of coupling, as the psychologists (and philosophers before them) had postulated.

6. Association

The principle of association inherent in this has aspects related to logic.

It illuminates the relation between *induction* and *deduction*. If all the roses anybody ever saw were red, the cell assembly which represents roses in the brain will be so strongly connected with that representing the colour red that the corresponding concept will be "red rose" rather than "rose". This much is induction and is conditioned by experience. But the statement which will arise in the brain "all roses are red", when referred to the internal representations is really "all red roses are red" which is deductive. Thus we may see the brain as an apparatus which codes the input in such a way as to transform inductive inferences in deductions in the internal language of the brain.

There is the puzzle of the extraordinary historical success of *Aristotelian logic*, based on the apparently arbitrary selection of propositions "all A are B", "no A is B", "some A are B" and "some A are not B". Reasoning in terms of brain mechanisms, and particularly in terms of cell assemblies, this choice seems a rather natural one. It is not difficult to imagine a mechanism which makes use of an associative memory (such as we envisage in the cerebral cortex) in order to test whether a certain input configuration elicits the ignition of a cell assembly. If it does, the configuration is thereby established as representing a fact (or thing, or event) which is known to have occurred. The four elementary propositions of Aristotelian logic: a, universal affirmative; e, universal negative; i, particular affirmative; o, particular negative are thus easily obtained:

If AB ignites a cell assembly and A$\bar{B}$ does not: a
If A$\bar{B}$ ignites a cell assembly and AB does not: e
If AB ignites a cell assembly: i
If A$\bar{B}$ ignites a cell assembly: o

7. Terms

Finally, the structure of the cortex suggests an interesting interplay between the *terms* and the *rules* that govern their interaction, something which the science of logic has perhaps not yet considered. The dominant neuron types of the cerebral cortex, the pyramidal cells whose interconnections form the associative matrix in which cell-assemblies develop, (fig. 1) are characterized by their peculiar shape. Their dendritic tree (= the receiving part, black) is composed of two portions, the "apical" and the "basal" dendrites, and the axons (= the transmitting part) are also two-fold, consisting of a local and a distant ramification, the first connecting

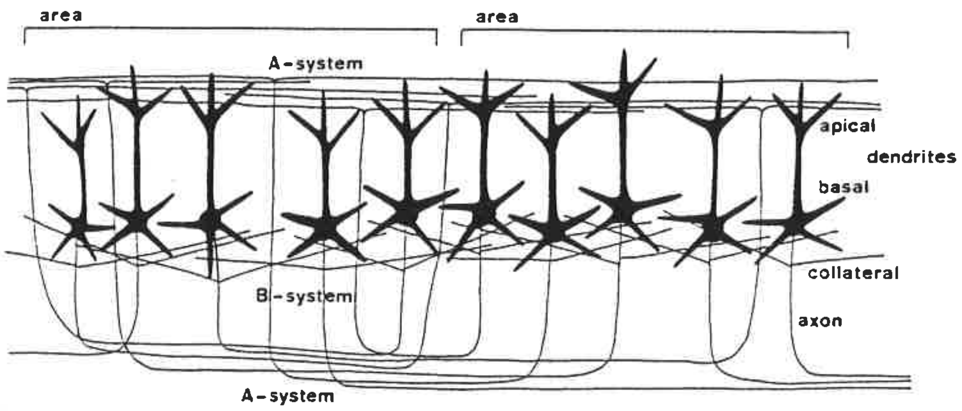

fig. 1

with the basal dendrites of nearby pyramidal cells (axon collaterals, B-system), the second with the apical dendrites of pyramidal cells anywhere in the cortex (A-system). The two systems of interconnection are about equally powerful, but they are different in one important respect: while in the B-system the probability of a contact between two pyramidal cells depends on their distance in the cortex (by some approximately inverse square or simply inverse proportionality), no such dependence is noted for the A-system, so that one could be called the metric system, the other the ametric one (terms suggested to me by Palm).

The cortex can be subdivided into a number of areas, defined by the special input they receive (acoustic visual etc) or by their role in generating the output (motor areas). Within each area there is spatial order, related to some sensory space or to the organization of the motor system. The spatial order does not generally carry over from one area to the next. We may say that within each area things happen in a uniform context, and in different areas in different contexts.

Now, the fibers of the B-system are mostly confined to one such area of uniform context, while those of the A-system mediate information between different contexts.

I offer the following interpretation. A thing or event of our experience is a bundle of properties in general pertaining to different contexts and held together by the fibers of the ametric, the A-system. On the contrary, movements or more generally evolutions of these things correspond to the regular changes of their various aspects represented separately in different areas of the cortex. We expect knowledge of these dynamic aspects to be represented in the narrow range fibers of the B-system. From the point of view of logic, the A-system represents the terms and the B-system the rules of their succession.

There is an important consequence of this. Since each cell pertains to one and to the other system, and since the elementary act of learning is most likely conditioned by the activity of the entire neuron, it seems that the learning of terms and the learning of the rules of their dynamics involve each other. This makes sense: terms are defined not only by their internal consistency (represented by a cell assembly via the A-system) but also by their usefulness in the discovery of a dynamics (represented by the synapses in the B-system). This would appear as a useful coupling of two different statistical aspects of the input, and certainly as a reasonable principle in coding the environment. The lesson we learn is that a rigid distinction of terms and rules may be misleading.

The theory of cell assemblies is appealing but far from experimental proof. However, if we see in the cell assemblies the physiological counterpart of the discrete elements of thought we have a plausible candidate for the puzzling discreteness which we observe at different levels in language. We must not forget that the discrete character of logic is inherited to a large extent from the discrete character of language.

References

Braitenberg, V.: Cortical architectonics: general and areal. In: Architectonics of the cerebral cortex. M.A.B. Brazier, H. Petsche (eds.), Raven Press, New York, pp 443-465 (1978)

Braitenberg, V.: Cell assemblies in the cerebral cortex. In: Lecture notes in biomathematics 21. Theoretical approaches to complex systems. R. Heim, G. Palm (eds.), Springer Verlag, Heidelberg, pp 171-188 (1978)

Braitenberg, V. and Schüz, A.: Anatomy of the cortex. Statistics and Geometry. Springer Verlag, Heidelberg (1991)

Hubel, D.H. and Wiesel, T.N.: Binocular interaction in striate cortex of kittens reared with artificial squint. Journal of Neurophysiology, vol. 28, pp 1041-1059 (1965)

McCulloch, W.S. and Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical Biophysics, vol. 5, pp 115-133 (1943)

Shannon, C.E. and McCarthy, J.: Automata studies. Princeton University Press (1956)

Wiesel, T.N. and Hubel D.H.: Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kitten. Journal of Neurophysiology, vol. 28, pp 1029-1040 (1965)