

A TOPOLOGICAL THEORY OF THE BRAIN

E.C. Zeeman

We use mathematics to try and explain the relationship between mind and brain, between memory and anatomy, and between thinking and the electrochemical activity of the cortex.

There is a tendency to be put off by the apparent complexity of the problem, but this is wrong because it reflects an overemphasis on the quantitative as opposed to the qualitative. In fact science has two faces, represented by the two columns (they are of course related) :

quantitative	qualitative
prediction	explanation
measurement	form
analysis	geometry

Although scientists pay lip-service to the left column, they are really more interested in the right. In physics observations often look so random (e.g. the motions of the planets) that physicists have to build elaborate mathematical models (gravity, calculus) in order to discover the simplicity of the underlying geometric form (ellipses). In some physical problems only the form and not the prediction is interesting. For example, the zig-zag motion of a falling piece of paper is a fascinating form, that can be handled mathematically, but the prediction of where on the floor the paper will land is too complicated to be of any interest. Meanwhile in biology, up to now, the forms have generally been so obvious, that qualitative and empirical observations have been quite adequate, so that very little mathematics has been needed.

But biology is now entering a new era, and it is beginning to need more sophisticated mathematics in order to discover the forms underlying the behaviour of complicated things like a brain. Analysis and differential equations are no good, because they are too precise and too local in character. We need a geometrical approach. We invent concepts such as tolerance spaces, and comparison spaces which are rather like topological spaces, and as a tool we use algebraic topology, because this is a brand of mathematics well adapted to ignore local variations and capture global properties. We want to capture global properties, because thinking is a global activity of the whole cortex, and we want to ignore local variations because thinking is unimpaired by the removal or loss of a few thousand neurons. We therefore build a topological model of the brain, which provides a very general framework, within which we can discuss a variety of phenomena.

continued ...

The main phenomenon that the theory tries to explain is that of visual perception; and the idea behind it is as follows. Consider a picture presented to the retina; the message travels down the optic nerve and the picture is preserved to a certain extent in the visual cortex, but the further the message penetrates into the cerebral cortex, the more the picture seems to be lost. Eventually one asks the question, "How does the brain ever reconstruct the message into a picture again?". But this question is a bad question because not only does it suggest that there is some sort of computer in the brain to do the reconstruction but, what is worse, it suggests that there is a 'real brain' somewhere at the back which contemplates the reconstructed picture. In fact we take the opposite view, namely that the message is never reconstructed into a picture, but is dissipated into a wave pattern over the whole cortex, which bears little resemblance to the original picture. We claim this does not matter; what matters - and this is what the theory attempts to prove - is that the set of all such patterns has a global mathematical structure similar to that of the set of all pictures. In other words the patterns behave just like pictures, and in fact are the pictures in our minds. Similarly, different colours cause patterns that behave like colours, and different sounds, through the auditory nerve, cause patterns that behave like sounds. The mathematical structures of the colour and sound patterns have different properties, and this explains why a mixture of two pure colours does not look like a chord sounds, for example.

The theory also explains how the local molecular changes inside the neurons and synapses can be combined mathematically to give something global which resembles what we think of as memory. An advantage of the theory is that it does not presuppose any particular mechanism of molecular change, but needs only the assumption that the local changes taking place are correlated in some way with the local activity. At the same time the theory generalises the circuit theory of memory. Just as the oceans at any given moment contain the superposition of the dying waves from all the storms in the last week, so the brain at any given moment contains the superposition of the dying wave patterns from inputs over the last 15 minutes. As each wave pattern fades it makes minute alterations in synapses all over the brain, and hence leaves a permanent "trace".

As an example of how the theory provides a useful framework for discussing different phenomena, we analyse certain skills possessing a rhythm or periodicity, such as walking or counting. It appears that the corresponding pattern of brain activity consists of a periodic control somewhere, possibly in the mid-brain, together with carefully phased delays to various parts of the cortex. The learning of the skill consists in making tiny alterations in all the synapses so as to set up the correct phasing of the delays. Recent experiments by Adey, Penfield and others suggest that this form of brain behaviour (periodicity plus delays) may be much more general than was previously suspected. It may underlie not only rhythmic skills, but all memory-sequences, as for example the memories of songs or how to solve problems. In this way the theory may help us to discover the essential geometric forms underlying the fantastic richness of our mental experience.