Simulating memory by numbers

The difficulties of making progress in neuroscience stem from the difficulty of formulating interesting (and answerable) questions. Should not associative networks be given more attention?

THERE is a sense in which neuroscience has been a great disappointment. Ten years ago, it seemed as if the question of how brains function was about to be answered by experimental investigation and analysis. There were new techniques for describing the anatomical connections of neurons and even for telling which of them were active at different times. Similarly, the catalogue of known neurotransmitter substances, already considerable, was being extended steadily while techniques such as that of 'patch-clamping' pieces of neuronal membrane seemed to offer a means of investigating the microscopic interaction between neurons and their environment in whatever degree of detail might be appropriate.

The promise of the new techniques has been sustained, yet people are not much the wiser about the working of the brain, no doubt for the familiar reason that it is easier to catalogue and describe its components than to know what questions most usefully to ask about their functioning. In the circumstances, neuroscientists should not be surprised if attempts at defining proper questions appear in unexpected places. Here is one intriguing notion, a scheme for explaining how a network of simulated neurons might be stimulated to regenerate not merely a specific pattern, likened to a memory, but a sequence of memories in predetermined order. An account of this trick, by H. Sompolinsky and I. Kanter, two physicists at the Bar-Ilan University in Israel, appeared in *Physical Review Letters* last month (**57**, 2861; 1986).

In reality, the new calculation is an extension of a train of rumination going back some years and most easily recognized by the label 'associative memory'. The objective is to design a network of physical entities each capable of existing in one of two distinct states (as neurons may be ON or OFF) which can simulate the behaviour of a real neural network. Those with long memories will recall that it is now thirty years since Professor Marvin Minsky launched his 'Perception', a conceptual network that might have made a working computer as well.

There are a few obvious analogies that promise that the search for a network model of the brain should not be fruitless. The hysteresis of ferromagnetic materials is a sign that strictly physical systems retain a memory of past states, while the materials known as spin-glasses, with a virtually infinite number of distinct ground states of similar energy, are plainly capable of greater sophistication in which each equilibrium state may in principle represent a different memory.

The usual formulation of the neuronal network problem is, indeed, formally identical with that of the problem of spin-glasses. Essentially it is the creation of J.J. Hopfield of the California Institute of Technology (see *Proc. natn. Acad. Sci. U.S.A.* **79**, 2554; 1982). Each node of the network is occupied by an ON/OFF entity, which in the case of a magnetic material would be physically a magnet pointing in one direction or the opposite. In principle, the neuron at each node is connected to all others in the network, which means that the properties of the network are determined by a set of coefficient T_{ij}, one for each pair of neurons i and j. If the first has no influence on the second, the corresponding coefficient is zero. If one neuron can influence another, the magnitude of the influence will be the product of the coefficient and a variable representing the state of the first neuron (mostly simply, 0 when OFF and 1 when ON).

To fix ideas, it is possible to think of the several influences on a single neuron as voltages that may be added algebraically (although more complicated schemes are permissible and, indeed, more realistic), in which case the target neuron will fire (or be ON) when the voltage exceeds some threshold. The state of each neuron is affected naturally by noise; Hopfield's model supposes that the state of each neuron at any time is nevertheless determined by a kind of instantaneous monitoring of the combination of all the influences that affect it. The neat conclusion of Hopfield's original calculation is that it is possible to choose the coefficients T_{ij} in such a way that a large number of arbitrarily chosen configurations of the system are stable against random fluctuations, or noise. The idea is that particular configurations may then be reliably evoked by suitable stimuli of the network as a whole.

Obviously the model has many attractive features. For one thing, for example, it allows that each memory state is a property of the network as a whole and not of any single element within it. For another, the model has the virtue of suggesting how a suitably configured network might be used to retrieve memories that are known, at the outset, only imperfectly; switch a subset of the neurons in the network to represent part of a memory recalled, and the result should be to lock the rest of the network into the corresponding entire memory. In this sense, it will be noted, the brain (if truly represented by such a network) is a good deal smarter than the microcomputer now on every other desk; 'files' are retrieved not by means of an artificial filename that must be accurately remembered but by part of their own content, which may be imperfectly remembered.

What Sampolinsky and Kanter have now done is to demonstrate that if the coefficients representing the connections between pairs of neurons are asymmetrical in the sense that T_{ij} and T_{ji} are not necessarily identical, it is possible to arrange that a Hopfield network will regenerate, in sequence, a predetermined set of its stable configurations. They find it necessary also to endow the neuronal connections (called synapses, of course) with time-dependent properties. The result is that the network, when it is stimulated, will run through a sequence of its stable configurations just as if its supposed owner were recollecting a movie of some kind.

What this implies is that neurons distributed through the network are turned ON and OFF in a manner determined by the sequence of numbered patterns to which they contribute positively or otherwise. Because the representation of memory cannot be exact, but will be influenced by noise, attempts to recognize successions of remembrances by recording from single cells would probably be doomed to failure. But the authors argue well that their model would probably be even more directly applicable to the working of those neuronal networks by which animals execute rhythmic movements, running or swimming for example. In short, ways of testing such models are probably not as inaccessible as they seem.

So far as they go, these calculations are good clean fun. By themselves, they do not suggest what questions might most profitably be asked of the working of the real brains of animals or people. But there is a case for asking whether these strictly conceptual models of neuronal networks have been given the attention they deserve by those who will eventually have to test their usefulness experimentally, psychologists and others. The results so far are almost too suggestive to be untrue.

John Maddox