Cognition

Parallel distributed processing

from Stuart Sutherland

THE human mind is sloppy and prone to error, but it has the saving grace that its mistakes are usually small ones. Until recently, computer models of intelligence exhibited none of these characteristics. They operated with great precision, and when corrupted, whether by a hardware or a software fault, gave rise not to small errors but to nonsense. Recently there has been much interest in a new way of simulating the mind that reproduces some of its sloppiness: it is called parallel distributed processing (PDP) or connectionism. One example of this type of processing is described by D.E. Rumelhart, G.E. Hinton and R.J. Williams on page 533 of this issue.

The PDP model postulates a set of units with properties similar to those of neurones. They are connected together with varying strengths, and are arranged in layers. Usually, there is an input and output layer and there may be intervening layers. In such systems the use of a given concept is represented by the activation in parallel of particular sets of units within each layer. A given unit may be active when very different concepts are entertained, though of course the pattern of active units differs from one concept to another. Because many units may take part in the representation of a single concept or mental process, small errors are likely to appear in the output as new learning changes the connectivity of some of the units. Again, because a single concept is not represented by a single unit, the system may continue to function after the loss of some units; in an analogous way, human memory often survives local damage to the brain tolerably well.

Such systems also generalize in a natural way. For example, if a system has learned to give a particular output to one item it will tend to give the same output to a new item that resembles the original one. Indeed, it has been shown that the more a new item resembles a prototypical item in a class, the more likely the system is to classify it as belonging to that class, just as people classify a robin as a bird more readily than they do a penguin. A further striking result is that the systems behave as though they are acting according to rules without the rules being explicitly represented, for the system's behaviour depends merely on the pattern of distributed connections within it. It is as though new instances were treated by analogy with old ones without any explicit rule being formed. This may well be the way children learn a grammar — they certainly have no conscious rule for determining what is the subject of a sentence and they may well have no unconscious representation of such a rule. Perhaps they merely treat new sentences as analogous to others they have encountered.

Apart from the psychological plausibility of PDP, it is clearly consistent with what is known of the nervous sytem. Moreover, its parallel architecture means that a great deal of processing can be achieved in a very short space of time. Synaptic delays and transmission time along axons restrict the brain to taking at most 50 serial steps in a quarter of a second, but this is long enough to recognize complex objects, an act that requires a vast amount of computation.

Quite apart from their naturalness, it has been found that PDP systems can simulate some of the more surprising findings about cognition. One system, for example, designed to learn how to form the past tense of a verb, makes much the same errors in the learning process as do children. Young children start by learning the commonest verbs, which are frequently irregular — "do, did"; "come, came". Children next master rarer and more regular verbs like "call, called". At this stage the child often begins to make mistakes with the irregular verbs it has already learned, saying for example "he camed" or "he comed". At the final stage, children become proficient in their use of both regular and irregular verbs. The computer model, although not designed to do so, mimics the children's performance at each stage. Several other PDP simulations have produced results remarkably similar to the behaviour of people. For example, a model simulating the learning of words came to distinguish between letter strings that conform to the pattern of English spelling but were not real words (for example, slet), and strings that could not possibly be words (for example, strz).

One of the problems in constructing such models has been that there was no adequate method of changing connectivities in a network with three or more layers. A possible solution is proposed in the report by Hinton and colleagues in this issue. The authors take the difference between the actual output and the desired output for a given input and uses a rule to change the strengths of the connections to the output in the direction needed to give the desired output. This procedure is then reiterated at each level of the system. He demonstrates that the recognition of symmetry and the learning of family trees can be achieved by this technique. One difficulty with learning systems of this sort is that optimal performance may never be reached because the system settles into a state in which the connectivity yields imperfect performance but in which any small changes to it only make performance worse (a local minimum). The learning rule of Hinton *et al.* yields good but not necessarily optimal levels of performance, but there is no proof that this would be true under all conditions.

Although for reasons not stated, they do not regard their technique as "a plausible model of learning in brains", it could be an important step forward. Whereas learning to associate two or more things together (for example, a person's face and his voice) can be achieved using PDP by existing learning rules that operate only at two levels, such rules do not apply to learning that depends on using the results of the outputs (knowledge of results): it is likely that here some form of back propagation must take place, and this is what Hinton and his colleagues have achieved. The problem is, however, often much more complex than are the authors' simulations. In a golf swing, the output is a series of muscle movements but the knowledge of results is too often the sight of one's ball being sliced into a ploughed field. How can this visual input gain access to the motor control system that produced the slice or change the strength of the connections that determined the swing, particularly when the exact strengths used on a given swing are probably no longer available by the time the ball is seen? The problem of the mechanisms by which knowledge of results, or even reinforcement, affects performance is an important one and has been largely overlooked by psychologists, and the ideas of Hinton *et al.* may provide a lead.

Regardless of this, the naturalness and power of PDP suggest that it is here to stay. Although some members of the old guard are resistant to these ideas, the doyen of the psychology of cognition, Professor George Miller, has remarked that it is the most important revolution in psychology in his day and his day includes the advent of cybernetics, information theory, generative grammar and the digital computer as a tool for simulating the mind. His pronouncement should not be taken lightly. Although there are many unsolved problems in PDP the existing results have been achieved by only a handful of workers. As the interest and importance of these systems become more widely known we may expect the young and flexible to explore their implications in detail, using no doubt the sloppy analogies that characterize both the human mind and PDP and that are almost certainly the foundation of creative thinking. □

Stuart Sutherland is Professor of Experimental Psychology at Sussex University, Brighton BN1 9QG, UK.