

Emotional arguments

Stuart Sutherland

The Muse in the Machine: Computers and Creative Thought. By David Gelernter. Fourth Estate/Free Press: 1994. Pp. 211. £16.99, \$22.95.

MOST workers in artificial intelligence are intellectuals. In consequence, most of them believe that if they can write a program that reasons and solves problems as well as they do, they will have created a replica of the human mind; some even claim that the program will be conscious. In trying to construct programs in their own image they have, however, lost sight of one of the most important features of the mind, namely emotion. The great merit of *The Muse in the Machine* is that it tries to put emotion into artificial intelligence; its great weakness is that it fails.

David Gelernter argues that human thought is a spectrum, ranging from focusing on solving a particular problem ("directed thinking" in the words of the Gestalt psychologists), down to the inconsequential thinking characteristic of dreaming, day-dreaming, the thoughts of young children and (more doubtfully) those of the ancients. His main claim is that this unfocused thinking is driven by emotion: successive thoughts are related only by their emotional overtones, not by any concrete similarity between their objects. He argues that it is such thinking that underlies all creativity and that, above all, gives rise to useful analogies.

It is of course true that emotion colours thought: those who are depressed think of depressing events and are pessimistic about the future. But to argue that all creativity depends on emotional linkages seems a gross exaggeration. Gelernter cites the well known case of Kekulé solving the problem of the structure of benzene by imagining a ring of snakes. But it is hard to see an emotional link between snakes and the benzene molecule: the metaphor surely arises from a structural similarity, not an emotional one.

Gelernter makes the point that creative insights often occur when the mind is relaxed — Kekulé was dozing in front of his fire when the image of the snakes came to him. There is, however, a perfectly good existing account of why creative images often occur when the mind is idling: it has been known for 60 years that under intense concentration only the strongest associations occur. For weaker ones to emerge, including analogies that are not immediately obvious, attention has to be relaxed. Not only is this phenomenon well supported by evidence, there is a plausible theory (going back to the underestimated Clarke Hull) to explain it.

Gelernter maintains that emotions are

based on bodily sensations, presumably, although he does not say so, mainly those from the autonomic system. But although this system is implicated in states of high arousal, it is most unlikely that its activation could account for all the subtle differences in human emotions that he stresses. Stanley Schachter showed that the arousal produced by adrenalin was interpreted as happiness by people in a happy situation and as anxiety by those in an unpleasant one. Moreover, someone moved by a poem may feel the appropriate emotion with no arousal of the peripheral nervous system. Nor can facial expressions or bodily posture account for emotion — "One may smile and smile and be a villain".

Gelernter describes a computer program he wrote to simulate directed thinking. In essence it superimposes images of the same kind of scene viewed repeatedly. The representation of anything that occurs on every occasion will be strengthened so that when the scene is evoked again, those features will be the most salient and the easiest to recover. He claims that syllogisms can be solved in this way, but as the program, unlike associative networks, is not supplied with the concept of class inclusion, it is hard to see how this can be done.

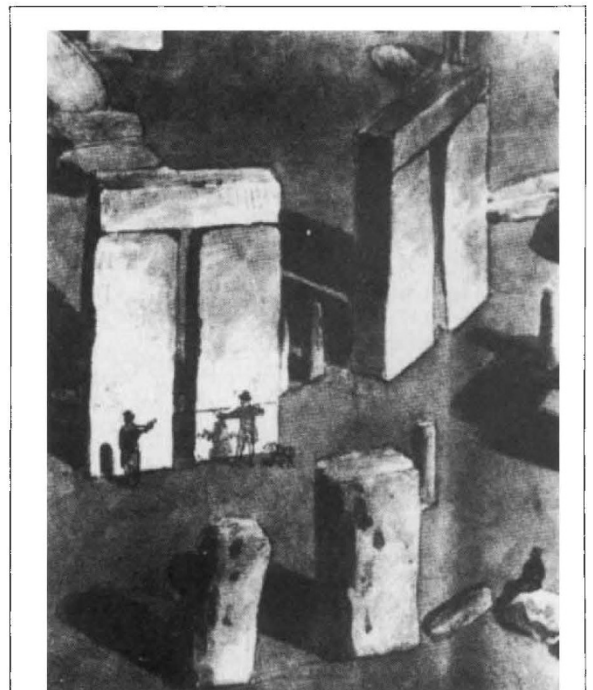
Oddly, and perhaps wisely, he has not attempted to write a program to simulate the role of emotion, although he believes this to be possible, by making the program move successively to representations that have the same emotional context but are in other ways completely different. He equivocates on whether such a computer would be conscious, eventually deciding that it would not be. He rightly observes that if someone says to a computer "I've lost my job" and the computer says "I understand", its reply is false, unless it can in some way feel the appropriate emotion. It may "understand" the facts, but not the blend of emotions engendered in the dismissed employee. It cannot empathize.

This point was made in

stronger terms by Joseph Weizenbaum many years ago. He pointed out that because of our biological nature we have drives such as hunger and sex, and the desire to love or be loved. We have these needs in order to survive and we are unhappy if they are unfulfilled. No computer could genuinely have them: at most it might exhibit a desire for the nearest electric plug. No matter how successful we were in simulating these emotions and drives on a computer, the program would still be a simulation, not the real thing. So there could be no reason to ascribe any form of consciousness to it.

Unfortunately Gelernter's writing and sometimes his thoughts tend to be rather jerky and he is given to annoying colloquialisms, such as "Nice try but no cigar" to describe an argument with which he disagrees. Moreover, he feels the need to support his own arguments by quoting banal remarks from such self-appointed sages as George Steiner ("Metaphor relates hitherto unrelated areas of experience") and Daniel Dennett ("Perception turns imperceptibly into memories"). Nevertheless his main thesis is aired too rarely and the book should be read by anyone interested in computers and the mind: it might even persuade some hackers to abandon their consoles from time to time and take a doze by the fireside. □

Stuart Sutherland is in the Laboratory for Experimental Psychology, University of Sussex, Falmer, Brighton BN1 9QG, UK.



MEASURING of Stonehenge early in the nineteenth century. Its design and proportions attracted the attention of architectural historians of the time. Taken from the revised and updated *Stonehenge Complete* by Christopher Chippendale. Thames and Hudson, £12.95 (pbk).