Strolling up the garden path

Igor Aleksander

Mind Children: The Future of Robot and Human Intelligence. By Hans Moravec. Harvard University Press: 1988. Pp. 214. \$18.95. To be published in Britain on 9 December, £14.95.

Mind Children is a frightening tale of the potential for the development of super-human 'intellect' in machines. The horror lies not so much in the actual content of the book, but in the fact that Hans Moravec suggests that a future in which the biological has been eradicated in favour of the technological is 'wonderful' and 'exciting'. To be fair, Moravec, as a sensible robot designer, gives himself trillions of years before seeing the realization of some of his predictions. But, he believes, this length of time can be bridged by 'transmigrating' a human being (consciousness, intellect, personality and all) into an immortal machine. Starting with the premise that in the near future there will be no essential human function, physical or mental, which cannot be artificially created, his argument follows a simple six-point progression.

Foundations

First, Moravec submits that the foundations of all the necessary techniques are already in place. Heading the list are Minsky's and Winograd's robot planning schemes, devised at the Massachusetts Institute of Technology in the late 1960s, which are said to be capable of understanding simple English-like sentences about worlds with simple objects in them. As further evidence, programs are said to exist that recognize sensory patterns and are capable of simple reasoning about them. Mobile robots of considerable agility have been designed, while learning and making general inferences has been shown to be possible with artificial neural networks. One problem, concedes the author, is that the technical knowledge has, as yet, not been married to engineering practice. This is needed to create a knowledge-gathering robot which can build up its own awareness of the world. Moravec accounts for this failing by suggesting that, because of the excessive cost, computer power has not reached the levels required. But given about 50 years, he says, the problem will have been overcome.

Consequently, the second pillar of Moravec's philosophy is that the acceleration in computer power per unit cost will continue. In particular, the ability of machines to absorb and classify masses of sensory data will increase hugely, and enable the mobile robot to explore its

surroundings in much the same way as a child. Moravec includes advances in biotechnology as a way of achieving some of the necessary packing densities of computing components.

The third step is to do with the symbiosis between man and machine. Improved sensors and displays will enable machines to work alongside human beings, allowing us to become helpful teachers to the robots. Devices such as 'magic glasses' and 'magic gloves' are invoked as existing early steps in making the computer provide sensory feedback to its human user. The gloves enable the user to 'feel' a simulated solid and so grasp and move objects on a screen. The glasses complete the illusion of actually being in some location that is merely stored in the memory of the machine.

The climax of this argument comes in the fourth step, where the world, driven by industrial competition, will see the development of the perfect computer receptacle into which the complete content of a human being's brain (hence intellect) can be transmitted. Moravec is aware that this notion can be demolished by the argument that such data would be useless to a machine that was not driven by its own needs to acquire it. He labels such non-believers as being besotted with some obscure viewpoint called 'the body-identity position', and urges them to shed this heresy and instead embrace 'the pattern-identity position'. This he clarifies as being the belief that 'I' am the process and pattern of what goes on in my head, and if that can be preserved in some other box 'I' become preserved in that box. In this way we shall all be able to be preserved forever, regenerating in a new box whenever the old one starts wearing out.

The fifth link in the progression suggests that a kind of independent, informational life can be created in the computer receptacles without the need to absorb human intellect. There is nothing to stop the receptacles from developing unlimited super intellects of their own. The evidence for this possibility is found in computer viruses (programs that destroy other programs unbeknown to the computer user); Conway's Life Game, which is played on a two-dimensional grid on which each point has simple rules about 'firing' or 'not firing' dependent on the firing of its nearest neighbours (limited patterns of grid activity survive and even shuffle around the grid to the surprise of the programmer); and Von Neumann's self-reproducing automata (programs transmitted from area to area on a grid of processors).

So we come to the sixth point, the *coup de grâce*, which defies the second law of thermodynamics. Because of the fall in noise with a fall in temperature, Moravec suggests that informational intelligence can remain constant as the Universe drifts towards absolute zero. What happens is

not clear, but by this point one does not care all that much.

Limitations

The trouble with predictions of this kind is that their persuasiveness lies in the disregard for the limitations of precisely those technological developments on which they are based. These deficits are hard to keep in mind as the reader becomes progressively blindfolded and is taken up the author's garden path. I would recommend that anyone who reads this book should immediately read another: Terry Winograd's and Fernando Flores's *Understanding Computers and Cognition* (Addison Wesley, 1986). This is a closely argued antithesis of Moravec's view, in which Winograd and Flores point out that 'intelligence' and 'understanding' as currently used in the context of computers may simply be metaphors that have no relevance to the creation of artificial intellect. From this standpoint, Moravec's reasoning collapses like a house of cards.

The ability of computers to 'understand' is shown to be merely a logical party trick (a juggling of pre-programmed phrases) by the very author (Winograd) of the programs on which Moravec bases the first step in his chain of reasoning.

Further, computer speed, cost and packing density have no bearing on machines becoming human-like. The missing ingredient is human consciousness itself, which is an integration of the experience of an individual's brain and body, closely tied to that body's make-up. This puts paid to Moravec's second and third pillars as it makes the 'body-identity position' seem sensible, and relegates the 'pattern-identity position' to science fiction. The rest falls into a crumpled heap: in particular, the Life Game and self-reproducing automata return to the realm of mathematical diversions, as relevant to philosophy as winning at Monopoly would be to becoming a real millionaire.

Moravec's book should be read, but perhaps not for the reason it was written. Where the author wishes to impress his audience with a vision of the way computer technology might lead to human immortality, readers should enjoy using no more than common sense to see if they can spot where the line between futures that can reasonably be predicted and those that belong to science fiction can be drawn. By not acknowledging the existence of views which differ from his own, Moravec makes no such judgements. For someone who is described as a serious scholar on the jacket of the book, that is a serious failing and one which vitiates his entire enterprise. □

Igor Aleksander is Professor of Neural Systems Engineering and Head of the Department of Electrical Engineering at Imperial College, University of London, Exhibition Road, London SW7 2BT, UK