Thinking about thought

Douglas Hofstadter

The Creative Mind: Myths and Mechanisms. By Margaret A. Boden. *Weidenfeld & Nicolson/Basic Books: 1990. Pp.303. £25, $24.95.*

I HAVE long admired Margaret Boden's style and mode of synthesis. Her 1977 book *Artificial Intelligence and Natural Man* (Basic Books) worked marvellously both as a lay-level introduction to artificial intelligence (AI) and as a scholarly overview. A main theme there was that AI can be seen as a quest for the true complexity and depth of the source of our humanity — our minds.

In her new book, Boden focuses on creativity: the most exalted, most human property of the mind. She convincingly argues that creativity is an emergent consequence of billions of coordinated micro-actions in a physical substrate (the brain); that creativity therefore can, at least in principle, be approximated on a computer; and yet that this view need not make us fear being reduced to a bunch of empty, meaningless circuitry. On the contrary, Boden argues, the mechanisms revealed ought to inspire awe at the subtlety and intricacy of the stuff of which we are made.

The strongest aspect of *The Creative Mind* is its discussion of human creativity, rich in insights about the role played by self-imposed constraints, the role of randomness, and creativity's connection with perception and memory. The weakest aspect, in my opinion, is the discussion of several highly vaunted computer models of particular mental processes, in which it is argued that each exemplifies a successful mechanization of some crucial element of creativity, even if on a very small scale.

One of Boden's deepest insights is that creative minds not only work under constraints, but in fact need them: "far from being the antithesis of creativity, constraints on thinking are what make it possible." She gives many examples, including grammar in language and tonal harmony in music. In discussing what might distinguish Mozart from ordinary mortals, she says: "These rare individuals . . . can search — and transform — high-level spaces much larger and more complex than those explored by other people. They are in a sense more free than us, for they can generate possibilities that we cannot imagine. Yet they respect constraints more than we do, not less. Where we can do nothing, or at best mentally toss a coin, they are guided by powerful domain-relevant principles on to promising pathways which we cannot even see."

The notions of 'conceptual spaces' and 'levels' of conceptual spaces are crucial to Boden's vision of creativity. A central thesis is that, to the extent one has explicit maps of one's own mind, one acquires new powers of creativity, and that there are unlimited levels of such self-reflection. She describes provocative experiments demonstrating striking differences between very unself-reflective attempts at creativity by very young children, and far more deeply creative acts by more highly self-reflective children just a few years older.

Among many fascinating examples of human and computer creativity discussed is Greek geometer Pappus' proof that an isosceles triangle's base angles are equal. The straightforward proof involves a construction cutting the triangle into congruent halves, but Pappus' proof involves no construction. The trick is to consider the triangle's mirror image as another triangle, then to prove these two triangles congruent. This devilishly ingenious proof, rediscovered two millennia later by a computer program whose author had never seen it before, seems like a perfect validation of the idea of 'creative computers'. But by keenly dissecting the program and revealing the nature of the conceptual spaces it had at its disposal, Boden deflates the accomplishment, pointing out that, in contrast to Pappus, "it could not recognize the interest of the proof it produced".

I only wish she had done the same for many other programs she discusses. Not that I wish to see AI as a whole deflated, but because it is a field in which results are incredibly hard to judge, great care is needed. There is a well-known effect, called the 'Eliza effect' after Joseph Weizenbaum's famous psychotherapeutic program Eliza, whereby people tend to impute far more depth to an AI program's performance than is really there, simply because the program's output consists of symbols — familiar words or images — that, when humans use them, are rife with meaning, and so when computers use them, it is next to impossible to suppress the unconscious aura surrounding them. Not just the lay public is susceptible to this insidious effect; surprisingly, AI cognoscenti are as well.

Thus one has a celebrated team of researchers claiming their program has precisely simulated the thought processes by which Kepler found his laws of planetary motion; then, in view of the discrepancy of the lengths of time it took their program and Kepler to do so, they feel compelled to invent excuses for Kepler's slowness — he had to sleep, eat and so forth. Never mind that the computer was presented with exactly the relevant pieces of numerical data, exactly the proper set of mathematical tools, and nothing else — in total contrast with Kepler himself, who worked in a fog where the boundary between what was relevant and what was irrelevant to the problem of orbits was completely uncertain, in an era when the very notion of scientific truth was a glimmer in one's eye, and belief in an intrinsic connection of mathematics with physical law was a great intuitive leap. Yet Boden devotes pages to this group's work with scarcely a cautionary word, leading her readers inevitably to the belief that much of the essence of scientific thought is captured by this model. Not only does this harm the field of artificial intelligence by building up unrealistic expectations; it also makes scientific reasoning seem far simpler than it is.

Elsewhere, Boden extols a program that allegedly 'understands' how Socrates' ability to evoke new ideas in the mind of an apprentice is like a midwife's role in a birth. When inspected up close, however, the program turns out to have no notion of midwifery, birth, 'ideas', 'mind', or Socrates; it merely has two skeletal charts of logical-dependency relations among a dozen or so formulas made up of uninterpreted symbols, and it searches blindly until it discovers a satisfactory alignment of the parts of the two charts. It is up to humans to read meaning into this act of skeletal-chart mapping; it certainly has no intrinsic connection with the understanding necessary for recognizing this — or any — deep analogy. Yet Boden, naïvely echoing the program's authors, claims that this program really gets at what understanding literary metaphors is all about.

To be fair, she certainly understands what today's researchers are doing and why. But I feel that she too unsceptically propagates researchers' own descriptions of their programs' achievements. Of course, too cynical an attitude would undermine her theme that important progress is being made in computer modelling of creativity. There is a middle ground, however: yes, the programs she touts provide some important new ideas about what minds are, but they do so collectively rather than individually. That is, although each program probably represents far less than its authors claim, nonetheless, when taken together, AI models do add up to a significant new viewpoint about the mind.

This book is undoubtedly not Boden's last word on minds and machines. She is committed to thoughtfully analyzing thought, and is one of the world's best commentators on these matters. Although I like the basic message of *The Creative Mind*, I hope she will go more deeply into these topics in the future. □

Douglas Hofstadter is at the Center for Research on Concepts and Cognition, Indiana University, 510 North Fess Street, Bloomington, Indiana 47408, USA.