Shadows of doubt

P. W. Anderson

Shadows of the Mind: A Search for the Missing Science of Consciousness. By Roger Penrose. *Oxford University Press: 1994. Pp. 457. £16.99, $25.*

ABOUT 15 years ago, Roger Penrose's former student, Stephen Hawking, devoted part of his inaugural lecture as Lucasian professor in the University of Cambridge to the prediction that by the year 2000 physicists would have been made obsolete by electronic computers. Although Penrose does not refer to this particular instance, it would seem that it is his concern about this kind of optimistic view of the capabilities of computers that motivated him to write *The Emperor's New Mind* and now, some five years later, its sequel.

In *Shadows of the Mind* he elaborates on his earlier proposals and attempts to answer his critics. I was put off reading *The Emperor's New Mind* by the many critical reviews it received, so I came to the sequel fresh, albeit prejudiced. Let me say without hesitation that my prejudices have been amply confirmed. Nonetheless, reading this new book is a fascinating and mind-stretching exercise. I can imagine that the average scientific reader, unfamiliar with the many-faceted mental universe that Penrose inhabits, will be dazzled by his extraordinary breadth and scope. But the more extraordinary the mind, the more unfortunate it is when it is used to entertain what may well be vain speculations. Also, Penrose's great reputation, charm and skill as a writer (perhaps not as evident in this book as in the previous one) should not blind us to the fact that his professional background is not really relevant to his subject matter.

The book consists of two parts. In the first part he argues that the mind does things that are beyond the capabilities of a "mere" computing machine. (The word "mere" is a trap; in this case it is in the meaning of "mere" that the meat of this statement lies.) This is why machines are not about to replace physicists (or mathematicians). I heartily agree. But he then concludes that some novel laws of physics must be crucial to the operation of our brain, and that they might possibly relate to certain aspects of quantum gravity theory. This conclusion troubles me. In the second part of the book, Penrose goes on to discuss his view of the gaps in our understanding of physics and biology through which such radically new material could creep into the theory of the brain. I can find little here to sympathize with.

He presents four alternative propositions about the mind that are intended to cover all possibilities: *A* is the "mere" machine; *B* is the machine with an impotent "Des Cartean" observer, which as far as its method of operation is concerned is essentially *A*; *C* uses possible new laws of physics; and *D* is the supernatural alternative that does not obey natural law. Penrose rejects *D* as a cop-out, as an alternative inappropriate to a scientist. *A* and presumably *B* are excluded by very subtle and ingenious reasoning involving a restatement of Gödel's theorem as an argument about computer algorithms: that no "provably sound" algorithm operating on a Turing machine or equivalent computer could ever encompass all the correct mathematics of which the human brain is capable.

To my mind, the most likely alternative is different from all of these; it is that the operation of the mind follows the ordinary laws of physics and chemistry, without

SPL

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Consciousness explained? Cross-section of a flagellum, showing its internal arrangement of microtubules.

bizarre additions, but that it operates using algorithms, concepts and mechanisms that are quite outside the system of apparently rigorous 'theorems' of computer theory. In computer complexity theory, for instance, the complexity classes are often meaningless categories. But by using one's knowledge of the nature of the problem to be solved, one can often do what from the theory seems to be impossible, whereas nominally 'equivalent' problems turn out to be inaccessibly distant from each other.

There are many ways in which computational methods using quite ordinary physics might evade the apparently 'rigorous' limitations of the von Neumann/Turing architectures. What follows are just a few, mainly culled from various recent books and articles about the mind.

First, the mind's hardware is by no means complete at birth. Not only its instruction kit, but its internal and external connections are constructed using knowledge of the nature of the actual world it will function in. The concepts of space-time and of objects moving continuously in space are built in; in fact, the most obvious optical illusions involve replacing discontinuous events with continuous ones. And most of all, the mind experiences objects: complexes of data that exist in space-time and show different aspects of the same entity.

Second, the brain's connections and its program are not complete until after it knows that other autonomous entities — other similar machines — constantly surround and communicate with it. There is a hint: the primitive mind animates with purpose even those objects in its surroundings that are inanimate. Communication is impossible without two factors: someone to communicate with and a common perception to communicate. Communication is a primary feature of mind, which is currently thought to be already established before the mind is complete.

Third, there is a fair amount of evidence that the mind is not a single, simple entity: it may be a number of independent, autonomous systems squabbling for a central dais. Multiple personality disorder is only an extreme form of what goes on in the mind all the time. There is no single Turing machine or single tape. It is not clear that it really is correct to model a parallel collection of semi-independent machines that is, in some sense, wider than it is deep, in terms of a sequentially operating single algorithm. In discussing complexity, this can be a different 'large-N limit', with different capabilities.

Fourth, some of Penrose's arguments, and much of computer theory, are about exact, rigorous solutions. His computers do not 'halt' until they have found an exact answer. This can be crippling. In the real world it is usually adequate to 'satisfice', to use Herb Simon's term. Methods directed merely at finding an acceptable way to do something can be much more efficient than exact ones. This is one way the mind can take advantage of its knowledge of the structure of the world.

As I see it, it is not really necessary to identify what particular aspects of the nervous system allow it to evade the rigid, rigorous, logical arguments with which Penrose tries to pin it down. One has merely to point to the remarkable ability of complex systems to develop emergent properties that overcome the apparent limitations of their separate constituents. Apparently rigorous 'theorems' that seemed to make antiferromagnetism as well as superconductivity impossible turned out to be irrelevant in the face of the emergent property of broken symmetry, just as all the many kinds of argument against evolution — the thermodynamic one, for example—do not prevent its happening. What does seem clear is that the above, and other new concepts and methods using conventional physics and chemistry, are far more likely to solve the problem of mind than is quantum gravity.

It is impossible to analyse here in detail all of the arguments in the second part of

the book. Let me pick out a few about which I have some independent knowledge. The long section on quantum measurement theory emphasizes the many dilemmas and queries that one encounters if one assumes, with Bohr, that there is a genuine dichotomy between the microscopic world in which quantum theory applies and the macroscopic world of measurement apparatus. These mind-bending difficulties ('EPR', 'entanglement', Bell's theorem and so on) are the stuff of rather boring philosophical discussions; it is hard to see how they could make consciousness easier to understand. But one seems unable to find any natural scale for this dichotomy, among other things; and many, if not most, thinking quantum physicists reject the idea that there is any dichotomy, and assume that quantum laws hold all the way up and down. This possibility is dismissed by Penrose in two brief pages (pp. 310–312).

Penrose's primary objection to this point of view is that it is "unsatisfactory" in that it involves continual splitting of the wave function of the Universe into fragments, only one of which an observer can perceive. (This splitting is the 'many-worlds' viewpoint, although there are other ways to interpret the same mathematics, among them that of M. Gell-Mann and J. Hartle.) We cannot decide for nature which of her ways are 'satisfactory' or 'unsatisfactory'; that is nature's call.

More seriously, Penrose makes the claim that there is no quantitative justification for the all-quantum viewpoint. In a popular book, *The Quark and the Jaguar*, published earlier this year, as well as in several articles, Gell-Mann discusses at length the rapid and complete 'decoherence' between alternatives, which prevents the observation of coexistence within, for a typical case, 10^{-21} seconds, by actual and precise calculation. That this is a consistent and logically satisfactory possibility has been obvious for many years, since. Fritz London first proposed it in 1938 It has now been formalized. Penrose should have been aware of this.

With regard to superconductivity, Penrose has, I think, got the implications of macroscopic quantum coherence backwards. In a superconductor, the quantum field itself becomes a macroscopic object, a perfectly measurable, rigid, thermodynamic parameter of the body on the same footing as strain, torque, entropy or magnetization, and obeying the same general laws (which derive from the general phenomenon of broken symmetry). Coherence is maintained not by an energy gap as Penrose suggests, or by some mysterious persistence of a quantum superposition, but by mundane thermal equilibrium. It has always seemed to me that for anyone in possession of the facts about superfluidity and superconductivity, it would be hard to doubt that classical behaviour is simply large-scale quantum behaviour — that is, an emergent property of large quantum systems. But habits of thought die hard.

Microtubules are for Penrose the likely seat of the mysterious quantum gravitational effect that makes the mind possible. Biophysicists who specialize in their study would agree that the behaviour of microtubules is indeed interesting and complex, but would see no need (nor in fact any room) for anything but the characteristic chemical control mechanisms with which we are familiar.

Penrose has written a complex, erudite and fascinating book, and my complaints about it do not mean that I did not enjoy and learn a great deal from reading it. But one should keep in mind that Penrose is a mathematician with little experience of the messy, frustrating but ultimately deeply satisfying process of checking his ideas against the experimental facts about nature. Mathematicians are used to game-playing according to a set of rules they lay down in advance, despite the fact that nature always writes her own. One acquires a great deal of humility by experiencing the real wiliness of nature. □

P. W. Anderson is in Department of Physics, Jadwin Hall, Princeton University, Princeton, New Jersey 08544, USA.

Mental pictures on the brain

Zenon Pylyshyn

Image and Brain: The Resolution of the Imagery Debate. By Stephen M. Kosslyn. *MIT Press: 1994. Pp. 516. $45, £40.50.*

THIS book is intended first and foremost as the final solution to the so-called 'imagery debate'. Its focus on this polemical task, however, seriously detracts from its potential usefulness as a study of the relation between visualization and vision, particularly from the perspective of clinical neurology.

Many of us believed that the debate, at least in the form revived in this book, had quietly disappeared as it became clear that there were serious problems with notions such as that mental images 'depict' or 'resemble' something or 'have spatial properties'. But Stephen Kosslyn now feels that these notions can be rehabilitated because "by turning to the brain, this debate can be resolved to the satisfaction of most people". But the basic problem still stands: as long as the research questions continue to be ill-posed, the problem about mental images will remain unsolved, regardless of how much brain (or other) data is collected.

Discussions of the nature of mental imagery have invariably equivocated between two very different views of what an image is. The literal option is that an image is some sort of mapping (usually viewed as a quasi-photographic projection) of the imagined scene onto some real (presumably neural) surface, possessing such physical and/or geometrical properties as shape, length, area and size. I do not know anyone who explicitly endorses this literal 'picture' option. In a way this is too bad because it is the only option that actually addresses much of Kosslyn's data (such as the increased time it takes to scan greater imagined distances or to examine smaller images). It is also the only option that clearly connects with most of the neurological findings discussed in the book. (Imagining something large, for example, results in brain activity over a larger area of cortex than imagining something small.)

The second option is that we have some 'functional equivalent' of pictures in our brain. Kosslyn talks about a "functional space" where images do not actually 'have' properties such as size or orientation but merely 'specify' them. But this option has no explanatory power because it fails to constrain the nonliteral 'image' to have any particular intrinsic properties — gone are depiction and resemblance, as are any constraints on how geometrical properties are represented. To account for empirical data one must of course reintroduce whatever additional constraints one needs, but these are no longer intrinsic properties of the image. As Kosslyn remarks, the critical properties are not inherent in the image but in how it is 'read'. Moreover, since such a functional image contains "previously digested information", there is no reason why 'reading' it should involve the visual system.

A good example of this sort of extrinsic stipulation of constraints is Kosslyn's use of a matrix as a functional image in his computer model. Notice that a matrix, by virtue of being a data structure, does not require scanning to proceed through adjacent cells, nor does it inherently preserve geometrical properties over transformations such as translation and rotation. Such constraints must be additionally stipulated. Consequently, appealing to the matrix itself does not explain predictions derived from such stipulated constraints, as it would if we had taken the literal option and assumed a surface constrained by the laws of physics.

The basic problem is that any theory of mental imagery has two fundamental degrees of freedom between which it can trade off in addressing the data, since the theory specifies both the nature of the image and the nature of the process that examines it. If we assume a literal view of the image, the physical geometry of the display allows us to make sense of some