Book Reviews

A Call for a New Physics

Shadows of the Mind. A Search for the Missing Science of Consciousness. ROGER PENROSE. Oxford University Press, New York, 1994. xvi, 457 pp., illus. $25 or £16.99.

In *Shadows of the Mind* celebrated Oxford physicist Roger Penrose continues, elaborates, and in some respects modifies arguments presented in his earlier book, *The Emperor's New Mind*. The claims put forward in these two extraordinary books are quite startling: that there are specific aspects of human experience, easily available to introspection, that cannot be explained, even in principle, within the framework of the known laws of physics; that in any case this framework is fundamentally flawed, in such a way that it fails in principle adequately to describe the behavior of macroscopic bodies; and that these two failings are related, so that only within a new physics, incorporating some yet-to-be-constructed quantum theory of gravity, will it be possible to understand the phenomenon of human consciousness.

Such claims, appearing as they do outside the regular scientific literature of the relevant disciplines, might not warrant serious consideration if they came from an anonymous source. But Penrose is hardly that, and his proposals deserve serious scrutiny. I have attempted such scrutiny, and—as will become obvious—have reached very negative conclusions. Before entering into the particulars, I should in fairness mention that large portions of this long book are devoted to exposition of background material and that this exposition is generally sound and occasionally brilliant (here I have in mind particularly the mini-biography of Cardano on pp. 249–256). Thus discriminating readers might benefit from the book without buying into its original scientific claims. My focus here, however, will be on the main line of development, which attempts to justify the claims sketched above.

This main line proceeds as follows. First, Penrose argues that humans can perform mental feats that simply cannot be performed by a machine, however complicated, that follows a finite algorithm (that is, a Turing machine). He observes that the conventional laws of physics, as they operate on a finite material system, can be simulated by a Turing machine. Thus he concludes that the mental feats of humans cannot be explained within the conventional laws of physics. Second, he argues that the conventional laws of physics contain the seeds of their own destruction, in that the conventional rules of quantum mechanics are logically incoherent and, carried far enough, must give incorrect results. This breakdown is supposed to occur for small but "semi-macroscopic" bodies (more on this below). Finally, he argues that the required new laws of physics, as applied in the human brain, will explain our ability to transcend the Turing paradigm. Let us examine each of these steps, in turn.

The central—indeed essentially the only—exhibit in Penrose's case that humans do things Turing machines cannot is the supposed "obvious" ability of human mathematicians to transcend the limitations of Gödel's theorem. The core of Gödel's argument is his construction, in any sufficiently powerful formalized system (roughly speaking, in any system specified that is powerful enough to deal with arithmetic and whose procedures could be completely specified and mechanized *à la* Turing), of a proposition that can be interpreted as stating "I am not provable." If this proposition is true, then it cannot be proved, and if it is false the system is inconsistent. Thus any powerful, consistent formal system will allow statements that are true but not provable. Now, says Penrose, we can see that any attempt to capture the power of human mathematicians by a formal system must fail. For a human mathematician could understand the *meaning* of Gödel's argument, even as applied to a hypothetical formal description of herself, and thereby recognize the *truth* of the Gödelian proposition. This, according to Penrose, demonstrates that the human has methods of reaching truth that have the force of proof but that cannot be captured by the proof-process of any formalized system. Experts will recognize (and Penrose acknowledges) the descent of this argument from a famous 1961 article by Oxford philosopher John Lucas, which spawned a large, contentious literature and certainly has not won universal acceptance. Among the many counterarguments that have been offered, one that appears particularly clear and convincing to the present reviewer is that the truth of the Gödel proposition only follows on the assumption of consistency—but consistency is, for a powerful system, not at all obvious. In fact, according to another closely related theorem of Gödel, consistency cannot be proved (if it is true). Thus the supposed royal road to truth involves a questionable shortcut, which arrives at something less than proof after all.

In any case, it seems quite strange to draw the battle line on this suspect terrain. Let us make a more modest request and ask for a demonstration that humans do something concrete that is not strictly impossible for Turing machines, but only difficult. And let us look where evolution suggests there might be something to find, in the perceptual processes that are likely (unlike mathematical logic) to be optimized by natural selection. Are there perceptual tasks that humans do much faster than any classical computer? (More precisely: Are there "holistic" perceptual tasks at the NP level of difficulty that humans can do in polynomial time?) Penrose seems to be edging toward this sort of question in his discussion of tiling problems, but he does not report on any attempts at systematic experimentation. Any convincing demonstration of such abilities, though giving only a much *weaker* result than Penrose claims, would have revolutionary implications.

Turning now to physics: Penrose perceives deep trouble in the foundations of quantum theory. He accepts that the physical interpretation of quantum theory *requires* something he calls the R-process, which is the moral equivalent of the "reduction of the state vector" invoked in some discussions of the Copenhagen interpretation. This is to be distinguished from ordinary deterministic dynamics described by the Schrödinger equation, which he calls the U-process. Some version of the R-process is, according to Penrose, necessary to make a bridge between the quantum rules for adding amplitudes and the classical rules for adding probabilities.

Again, this is suspect terrain. It is by no means the case that all informed physicists see the need for an R-process; indeed, the modern tendency (to which this reviewer is fully sympathetic) seems to be to see if we can get by with just the U-process. Though it would be very premature to declare final victory for this approach—that will require, I believe, construction of a recognizable caricature of an intelligent observer within the formalism of quantum theory, so that model experiences could be compared against our subjective experience as real observers—there have not been any decisive defeats, either, in some challenging battles.

What this means can to some extent be illustrated with reference to Penrose's ten-

Vignettes: Innovation Management

The American legal system's adjustment to the industrial revolution suggests that it will adjust to the technological revolutions that lie ahead. Thus we see again the contrast between the scientists' sense of a world making progress and the lawyers' sense of a more or less endless process of mediating social disputes. J. D. Watson exuberantly described his pathbreaking work in formulating the double helix model for DNA as "perhaps the most famous event in biology since Darwin's book." Yet in an early discussion of legal controls on recombinant DNA research, the prominent attorney and legal scholar Harold Green reported that he was "happy to say" that nothing "unique or novel" in such research insulated it from regulation.

—*Steven Goldberg*, in *Culture Clash: Law and Science in America* (New York University Press)

What are the duties of the inventor to his invention? To what extent must the inventor be responsible for seeing that his new invention and the world into which it is brought are compatible? Frankenstein's creation, while offering to do his duty to the doctor if the doctor will only fulfill his obligations and care for him, is nevertheless ready to exact revenge if the doctor does not. The monster's murder of Frankenstein's bride on their wedding night must be set against Frankenstein's refusal to make a companion for the monster, a refusal to help fashion an appropriate environment for his invention.

—*Cary Fowler*, in *Unnatural Selection: Technology, Politics, and Plant Evolution* (Gordon and Breach)

tative (and, it is fair to say, extremely vague) proposals. He claims that including the effects of gravity within the quantum-mechanical formalism will make a qualitative change in its predictions for the behavior of macroscopic bodies, in such a way that the laws of classical probability will apply for sufficiently large bodies. Well: what about the behavior of K-mesons, neutrons, and even photons, which are known by exquisite interferometric experiments to maintain quantum coherence after traversing what are by any reasonable measure macroscopic distances? What about superconductors, which conduct perfectly "through a mile of dirty wire" (Casimir), or similarly superfluid helium? What about the profound and beautiful work now being done on mesoscopic systems, which probes quantitatively how characteristic quantum mechanical behavior gradually becomes more subtle—but does not seem to break down or disappear—for pure, small, cold but definitely macroscopic systems (involving many thousands of electrons)?

The quantum theory of gravity is fraught with difficulties regarding its behavior at high energies and short distances, difficulties that may or may not be resolved by superstring theory. However, quantization of the low-energy, long-wavelength part of the Einstein theory is not problematic, and effects of the sort Penrose proposes ought to be discussed within this theoretical framework, as a first step. My own tentative conclusion is that the predicted effects are exceedingly small and are likely to be overwhelmed, under any remotely practical circumstances, by more mundane processes. In any case this framework is eminently *computable*, in the technical Turing-machine sense, so it is not sufficient for Penrose, who wants somehow to introduce a *noncomputable* R-process.

For the next step in Penrose's synthesis is to invoke the hypothetical noncomputable R-process to explain the supposed noncomputable abilities of human brains. Regarding this, I (like Penrose) will be brief. He claims, inspired in part by ideas of Hameroff, that microtubules perform crucial information-processing roles in the human brain and that they behave in essentially quantum-mechanical ways that allow them to transcend the limitations of Turing machines. A lot is known about information processing in the brain, particularly in the early stages of visual processing, and as far as I know it has never proved necessary to assign an important role to microtubules. Microtubules are not particularly characteristic of the human nervous system—indeed they are common in single-cell organisms—and on the face of it appear to be versatile structural elements in many classes of cells. Moreover, the conditions of heterogeneity and temperature characteristic of biological activity hardly seem conducive to quantum coherence on a macroscopic scale. So speculations about a spectacular computational ability of microtubules based on quantum coherence and central to human consciousness appear quite bold at this time. They, at least, would seem to be open to experimental investigation in the near future.

It appears to me, in summary, that Penrose's argument, from formal logic and philosophy, that human beings perform noncomputable operations is simply mistaken; that his argument that quantum theory is incomplete is unconvincing and his proposed remedy implausible; that his conclusion that an essentially classical description of microtubule function must fail is premature to say the least; and that his discussion of this topic, and of neurobiology in general, does not do justice to a large important body of relevant empirical knowledge. Moreover, the whole grand structure of his arguments is exceedingly fragile, being at no point buttressed by specific reference to nontrivial experimental facts. Perhaps not since the heyday of the great rationalist metaphysicians—Descartes, Leibniz, Spinoza—has there been a comparable performance. Although there are several brilliant passages in the book and the distinguished author is evidently sincere in his convictions, in the end one can only agree with Francis Crick, who wrote (commenting on *The Emperor's New Mind*), "It will be remarkable if his main idea turns out to be true."

Frank Wilczek
Institute for Advanced Study, Princeton, NJ 08540, USA

Cats in Groups

Cheetahs of the Serengeti Plains. Group Living in an Asocial Species. T. M. CARO. University of Chicago Press, Chicago, 1994. xxii, 478 pp., illus. $70 or £55.95; paper, $26.95 or £21.50. Wildlife Behavior and Ecology.

This book continues the tradition of excellent studies of the life histories of individual species that have emanated from the Serengeti, one of the world's most spectacular ecosystems. Apart from providing a synthetic account of the biology of cheetahs in the wild, Caro explores the evolutionary causes of social behavior and examines the behavior of cheetahs from a functional perspective.

This book contains a wealth of valuable and interesting data that have been carefully collected, expertly analyzed, and clearly