Physics, Mathematics, and Minds

The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics. ROGER PENROSE. Oxford University Press, New York, 1989. xiv, 466 pp., illus. $24.95.

Among the great variety of serious books on science, only a few focus on what we do *not* know. Even fewer succeed as true contributions to knowledge by anticipating problems that turn out to play a pivotal role in subsequent developments. The best of them can alter the course of science by changing the perception of what is and what is not a scientific problem. To what extent they are successful can be judged only with the benefit of hindsight. My bet is, nevertheless, that Penrose's best seller with its focus on the relationship between the mind, the universe it inhabits and investigates (physics), and the tools used in the investigation (mathematics) has aimed at the right target and will leave a lasting impact, even though I disagree with many of the answers Penrose proposes and am not convinced that the specific questions he suggests will remain at the center of attention.

"Can a computer have a mind?" is the central issue posed and discussed in the book. There is of course little doubt that a sufficiently powerful computer equipped with a sophisticated program can successfully emulate many of the "algorithmic" functions of the human mind: Computers are better than humans at checkers and better than all but a few humans at chess. Penrose recognizes all this, but builds a strong—if to this reader not entirely convincing—case against the computer's having a *real* mind. The nonalgorithmic (and hence impossible-to-program) functions are to include more "elusive" aspects of mental processes such as "understanding," "self-awareness," or "insight," which we know well from introspection but which are much harder to define operationally. Penrose argues that while we should continue to attribute them to fellow human beings, they should not be attributed to "mere programs."

The notion that natural intelligence is superior to the artificial one in some elusive but essential way has a great romantic appeal. If correct, it would require the laws of physics relevant to the operation of human brains to be essentially nonalgorithmic. For if they were algorithmic one could use them, at least in principle, to write a program that simulates a specific brain on a sufficiently fundamental (molecular? atomic? subatomic?) level. Hence, Penrose suggests, there had better be room for a nonalgorithmic ingredient in the law of physics.

What follows is a grand tour of the relevant parts of the theory of computation (intended to define carefully what is and what is not algorithmic), mathematics (to capture its nature and to define its role in the formulation of physical laws), and, above all, physics (in the search for the nonalgorithmic "holy grail"). Major vistas include Gödel's undecidability, fractals, classical and quantum physics, cosmology, the "arrow of time," and black holes.

Though the nonalgorithmic ingredient is not discovered in any of the known laws of physics and has to be postulated, the tour is a resounding success: The guide does not use the standard guidebooks, has direct research experience with a majority of the subjects discussed, has very strong opinions, and is unafraid to be controversial. For a casual reader, the tour will be an exciting (and challenging) introduction to science. For a practitioner, it will offer a fresh point of view.

A few words of caution: The book is not a casual "popular science." It will in parts be quite demanding of its readers (pleasantly surprising in a book that has become a best seller). It contains a fair number of equations—although they are employed as illustrations, to supplement rather than substitute for explanations. (An example: a two-page binary sequence with a Gödel number of a universal Turing machine!) Nevertheless, I do not know of any introduction to the discussed "frontier territories" that—especially for a reader of *Science*—is at the same time more accessible, as exciting, and yet equally deep.

The greatest weakness of the scientific journey that makes up the core of the book is perhaps inseparable from its greatest strength—the personal nature of the account. Many of the topics are presented in a very subjective fashion, clearly tailored to support the main idea.

For example, the section devoted to the various attitudes toward the measurement problem in quantum theory (exemplified in the book by the infamous "Schrödinger's cat") gives a brief description of both the "many worlds" interpretation due to Everett, according to which each measurement results in the wave function of the universe splitting into various "branches," each corresponding to a distinct outcome, and of the "environmental" point of view, according to which only one of the alternative outcomes is perceived because the coherent superposition between them is destroyed by the coupling with the environment. Both of these viewpoints are found lacking by Penrose. He argues (correctly) that the Everett interpretation alone does not explain why our (observers') consciousness perceives only individual branches of the "universal" wave function. He also notes that the viewpoint focusing on the environment and decoherence does not really account for the potential outcomes that do not "happen" when the measurement is made.

An obvious solution to this dilemma, not explored by Penrose but investigated by a number of others (including Caldeira and Leggett, Gell-Mann and Hartle, Joos, Unruh, Wigner, Zeh, and this reviewer), is to acknowledge that the environment influences directly the choices of the set of possible alternatives, classical states of physical systems, including the alternative states of the brain and therefore, by extension, "states of mind" as well. This position is perfectly in accord with the brain's being just a sophisticated parallel supercomputer, a complex (but "algorithmic") collection of interacting logical elements—neurons. It explains why quantum superpositions of classically incompatible states of mind are ruled out—they simply "decohere." It also lays to rest the old questions (which Penrose implies are still open) about the linear superpositions of dead and alive cats or of cricket balls at several locations and shows why the moon is not "washed out" into a coherent quantum superposition in its orbit as is an electron in the atom described by the Schrödinger equation, but rather rises and sets in accordance with the predictions of classical mechanics.

Whether one accepts the calculations that support this "decoherence" point of view as a stringent constraint on the possible solution of the dilemmas posed by the interpretation of quantum theory or only as a prescription for the determination of the "branches" in the Everett many-worlds interpretation is a separate question. The key point is that present-day understanding leaves considerably less room for maneuver than Penrose appears to imply. As a result, the book stops somewhat short of stating the real problem, which—at least for those of us unprepared to wholly embrace many worlds as the final answer—persists regardless of whether one is prepared to agree with Penrose on his main thesis about the elusive nature of human consciousness.

A reader who finds the above paragraphs

hard to follow is urged to reread them after going through the relevant sections of the book, which in spite of several similar (subtle) omissions and a number of unorthodox points of view is a superb introduction to what is most fascinating in science. The controversial aspects of the discussion present little danger to the practitioner and, though a more balanced account would have been even more useful, especially to the general readership, the author usually provides appropriate warnings so that nonexperts will not be led too far astray.

In a sense, the construction of the book is much like that of a detective novel, where the "crime" (the nonalgorithmic nature of consciousness) is identified early on (if only on the basis of somewhat circumstantial evidence) and various "suspects"—laws governing areas of physics potentially relevant to the operation of the brain—are introduced, thoroughly "interrogated," and found innocent. In the last sections the book returns to the "scene of the crime"—the human nervous system—and the detective (Penrose) points a finger at the presumed culprit.

A review of a detective novel should not spoil readers' fun by disclosing "whodunit," and I am not about to violate this rule. However, whereas the success of Poirot in Agatha Christie's novels is usually confirmed by a confession from the perpetrator, Penrose does not claim to be able to extract such an unequivocal "admission of guilt" from his suspect. This is just as well. I do not think it takes anything away from the excitement of the investigation. Indeed, in a sense it appears to be an open invitation for the readers to join in the on-going case.

The value of this book should be judged not just by the exciting overview of chosen areas of science, but, above all, by the fact that it puts into the center of natural science questions that so far have been asked mainly by philosophers and children. Penrose's book, I believe, anticipates the age in which science will have to come to terms with the fact that the minds that investigate the universe are inextricably embedded in its physics and in which the division between "mind" and "matter" will have to be either drawn more clearly or abolished altogether. When that happens, science in general, and physics in particular, will cease to be just a description of the universe by passive "detached" observers and, instead, will become a study of how minds are molded by matter and what role they play in the unfolding history of the universe they inhabit.

WOJCIECH H. ZUREK
Theoretical Division,
Los Alamos National Laboratory,
Los Alamos, NM 87545

Russians on the Psyche

Russian Psychology. A Critical History. DAVID JORAVSKY. Basil Blackwell, Cambridge, MA, 1989. xxii, 583 pp. + plates. $34.95.

David Joravsky, one of the leading historians of Soviet science and culture, has a fascinating story to tell in this book. Or, to be more exact, he has many fascinating stories—about Sechenov and the birth of Russian neurophysiology and psychology in the second half of the 19th century, about Pavlov, his career under Tsarist and Soviet regimes and the remarkable triumph of Pavlovism in the Stalin period, about the influence of Freud in Russia, about Vygotsky and his school of psychology in the 1920s, and (in an absorbing and provocative chapter that is really more an appendage to the book than an integral part of it) about psychiatry and political power in the Soviet Union from the 1920s to the 1970s. As the author freely admits, this is not a normal history of a science, focused on a single discipline and viewing it from essentially the same perspective as its practitioners. Joravsky's theme is the study of mind and brain in Russia. In other words, he is writing the history of two distinct and often competing scientific disciplines, neurophysiology and psychology. Far from being abashed by the duality of his subject matter, Joravsky is intrigued by it. Indeed, that duality, which he sees as symptomatic of a larger problem of "fracture and frustration" in modern culture, is an integral part of his theme; and it is the Russians' persistent but unsuccessful efforts to overcome it that compel his most serious attention.

"Starting in the time of Marx and Comte, of Dostoevsky and Tolstoy," Joravsky writes, "I ask how that old-time amplitude of spirit came down to Pavlov and his molecule of mind, the conditioned reflex." As this quotation suggests, Joravsky's approach to Soviet neurophysiology—and in particular to Pavlov, that "assertive little one-sided man," as he calls him—is not particularly sympathetic. But at least he concedes that neurophysiology is a legitimate scientific discipline with a real core subject and an accumulating body of knowledge. Not so for psychology, of which Joravsky writes that "the psychologists' findings have persistently failed to cohere within a cumulatively developing body of knowledge, or worse: different heaps of data have been diligently accumulated by different schools, only to sink into pointlessness as the schools go out of fashion and new ones win favor." This is a judgment of the discipline as a whole, but Joravsky certainly holds no special brief for its Russian practitioners, including those like Vygotsky and Luria, whose studies of child development and brain-damaged subjects are often admired in the West. There was "something in the science of psychology" (as well as something in the Soviet political climate of 1920s and '30s) that "restricted even the best minds to humble tasks of adjustment." In Joravsky's view, it is social scientists and humanists—"Marx and Comte, Dostoevsky and Tolstoy"—who have proved to be the best investigators of the human mind and psyche. Logically, given this premise and his subject matter, Joravsky's book includes quite detailed discussions of such efforts by Russian *non*-psychologists, including Tolstoy, Dostoevsky, Chekhov, the poets Tiutchev and Briusov, and the prose writers Isaac Babel and Iurii Olesha.

Still, it is science that is Joravsky's central concern in this book; and of the various threads of scientific development he follows, the longest and perhaps most colorful is that of Pavlov and the Pavlov school. Born in 1849, Ivan Pavlovich Pavlov was a distinguished and successful physiologist well before the revolution. Recipient of a Nobel Prize in 1904 for work on the digestive system of dogs, Pavlov subsequently developed the theory of conditioned reflexes (which, as Joravsky points out, should really be rendered in English as "conditional [*uslovnye*] reflexes), which appealed strongly to American behaviorists and led J. B. Watson to hail him as a master in his 1915 presidential address to the American Psychological Association. Pavlov was as uninterested in politics as he was in philosophy (the latter attribute being a major cause for Joravsky's distaste, as well as the subject of several irreverent and entertaining anecdotes in the book), but he had no initial sympathy for the Bolsheviks and objected to the scientific pretensions of their Marxist ideology. The Bolshevik leaders, however, respected Pavlov's achievements and international reputation and basically treated him well in the 1920s, providing his Institute with special rations and support and leaving even his hostile comments on Marxism unpunished, though not unrebuked.

In the late 1920s, Pavlov's work on conditioned reflexes had reached an impasse in scientific terms and was coming under seri-