A Review of Roger Penrose's THE EMPEROR'S NEW MIND: Concerning Computers, Minds, and the Laws of Physics

(Oxford University Press, 1989)

Harold G. Slovic

Assistant Professor of English, Aichi Gakusen University, College of Business Administration

＊The Emperor's New Mind is a fascinating, but difficult book. It is fascinating, because of the enormous range of topics which the author, Oxford University Professor of Mathematics, Roger Penrose, includes in the richly woven tapestry of his argument, from the uncommon–sensensical intricacies of quantum mechanics to the mind–warping speculations of contemporary cosmologists concerning the birth and ultimate fate of the universe. Along the way, Penrose brings to bear a number of subtle arguments taken from mathematical theory to support his views, and these are also, to those unfamiliar with them, interesting in their own right. But, it is a difficult book, due not only to the subject matter at hand, but also, unfortunately, in this winter's opinion, to certain decisions Penrose made in his manner of treating the material. I will attempt, in this brief review, to outline the main thrust of the author's argument, and to explain my reasons for feeling some slight degree of disappointment in the author's treatment of the subject matter.

A brief statement of Penrose's purpose in writing this book is required to give the reader a point of departure; in the Introduction to Chapter I, entitled "Can a computer have a mind ?", the author writes:

> The question of whether a mechanical device could ever be said to think—perhaps even to experience feelings or to have a mind—is not really a new one. But it has been given a new impetus, even an urgency, by the advent of modern computer technology. The question touches upon deep issues of philosophy. What does it mean to think or to feel ? What is a mind ? Do minds really exist ? Assuming that they do, to what extent are minds functionally dependent upon the physical structures with which they are associated ? Might minds be able to exist quite independently of such structures ? Or are they simply functionings of (appropriate kinds of) physical structure ? In any case, is it necessary that the relevant structures be biological in nature (brains) or might minds equally well be associated with pieces of electronic equipment ? Are minds subject to the laws of physics ? What, indeed, are the laws of physics ? These are among the issues I shall be attempting to address in this book. (pp. 3, 4)

Judging from a cursory reading of the above, there is little to indicate that this book will be any different from the hundreds (perhaps thousands !) of preceding books which have dealt with the seemingly unsolvable "mind/-body" problem—but, on second glance, it becomes clear that the issue here has as much to do with "machine intelligence" as it has to do with human beings—an issue the resolution of which has, according to Penrose, been given "a new impetus, even an urgency, by the advent of modern computer technology". And here, it is not so much the machines, in themselves, which have "raised the issue", but rather, a philosophical viewpoint held by some theorists who would claim that there is no difference, in principle, between the fundamental nature of "mind" as manifested by the human brain or as manifested in the workings of a mechanical device such as a computer—enter the "bad guy" of

the story, which Penrose refers to as *"strong AI":

> Thus, according to strong AI, the difference between the essential functioning of a human brain (including all its conscious manifestations) and that of a thermostet lies only in...much greater *complication*.... Most importantly, all mental qualities—thinking, feeling, intelligence, understanding, consciousness—are to be regarded, according to this view, merely as aspects of this complicated functioning; that is to say, they are features merely of the *algorithm* being carried out by the brain. (p. 17)

Penrose repeatedly gives the reader clear signals as to which side of the issue he stands on, as when he states, for example:

> In my opinion...a great many...people have been led astray by the computer people.... The belief seems to be widespread that, indeed, 'everything is a digital computer'. It is my intention, in this book, to try to show why, and perhaps how, this need not be the case. (p. 23)

To do so, Penrose makes reference to an astounding number of mathematical, physical, and philosophical ideas in an intricate weaving of argument which, I confess, often left my head swimming with disturbing feelings ranging from mild confusion to total incomprehension. The rubrics alone of the various (sets of) ideas Penrose makes use of give testimony to the involuted pathways of his argument, viz.: Turing machines. the Mandelbrot set, Hilbert space, Gödel's theorem, Shrödinger's cat paradox, etc., to name just a few. And the roster of names of the "heavyweight" thinkers in the fields of physics and mathematics whose ideas Penrose alludes to reads like a guestlist to some

improbable, spacetime defying convocation: Abu Ja'far Mohammed ibn Muse al–Khowarizm, Appolonios, Archimedes, Bohr, Boltzman, Cantor, Chandrasekhar, Dirac, Prince Louis de Broglie, Einstein, Escher, Eudoxos, Euler, Fermi, Feynman, Fourier, Gauss, Hadamard, Hawking, Heisenberg, Hofstadter, Hubble, Kepler, Lobachevski, Lorentz, Maxwell, Minkowski, Newton, Planck, Plato, Poincare, Ptolemy, Riemann, Russell, Rutherford, Schwarzchild, Wang, Wernicke, Wessel, Weyl, Wheeler, Whitehead, Wiesel, Wigner, Wilson, etc. (this is only about half of the list).

There are many things in this book which made me feel that my general knowledge of physics and mathematics is greatly inadequate, but one thing that truly surprised me was that Penrose puts Einstein's theories of relativity (the "special" and the "general") into the category of "classical theories", right alongside the ideas of Newton; evidently, the "worldshattering" ideas of Einstein, while still highly relevant and viable in providing a more "accurate" descriptive model of certain physical phenomona than Newtonian mechanics, are now considered to be somewhat "old hat", with the limelight now being focused on the newer theories of "quantum mechanics", the fundamental ideas of which were first developed during the time when Einstein was alive, but which have since undergone extensive elaboration. To help the reader understand the connection between the "classical" world and the "quantum" world, Penrose writes;

> In classical physics there is, in accordance with common sense, an objective world 'out there'. That world evolves in a clear and deterministic way. This is as true for the theories of Maxwell and Einstein as it is for the original Newtonian scheme. Physical reality is taken to exist independently of ourselves; and exactly how the classical world 'is' is not affected by how we might choose to look at it.

Moreover, our bodies and our brains are themselves to be part of that world. They, also, are viewed as evolving according to the same precise and deterministic classical equations. All our actions are to be fixed by these equations—no matter how we might feel that our conscious wills may be influencing how we behave. Such a picture appears to lie at background of the most serious philosophical arguments concerned with the nature of reality, of our conscious perceptions, and of our apparent free will. Some people might have an uncomfortable feeling that there should also be a role for *quantum theory*—that fundamental but disturbing scheme of thing which, in the first quarter of this century, arose out of observation of subtle discrepancies between the actual behaviour of the world and the descriptions of classical physics. To many, the term 'quantum theory' evokes merely some vague concept of an 'uncertainty principle', which, at the level of particles, atom or molecules, forbids precision in our descriptions and yields merely probabilistic behaviour. Actually, quantum descriptions *are* very precise, as we shall see, although radically different from the familiar classical ones. Moreover, we shall find, despite a common view to the contrary, that probabilities do *not* arise at the minute quantum level of particles, atoms, or molecules—those evolve *deterministically*—but, seemingly, via some mysterious larger—scale action connected with the emergence of a classical world that we can consciously percieve. We must try to understand this, and how quantum theory forces us to change our view of physical reality. (p. 225)

It is from this "mysterious lager—scale action" where the classical world evolves from the underlying strata of quantum events that Penrose hopes to

find one of his escape routes (he has others) out of the "deterministic" prison of the "strong AI" view of "mind". This is more easily said than done however, as, despite his lengthy and painstaking explanations of the fundamental concepts of the quantum view of physical reality, the average reader, who, like myself, lacks the appropriate mathematical background, is likely only to sense a vague glimmering of light, without actually being able to locate the point of exit within the maze of the author's argument and counter–argument.

In the chapter entitled "Quantum magic and quantum mystery", which occupies a major portion of the book, Penrose attempts to show how two different (mathematical) conceptual analyses are required to give a full description of the "evolution" of events from the quantum level of "reality" to the classical level:

> Regarding $*\psi$ as describing the 'reality' of the world, we have none of this indeterminism that is supposed to be a feature inherent in quantum theory–so long as ψ is governed by the deterministic Shrödinger evolution. Let us call this evolution process **U**. However, whenever we 'make a measurement', magnifying quantum effects to the classical level, we change the rules. Now we do not use **U**, but instead adopt the completely different procedure, which I refer to as **R**, of forming the squared moduli of quantum amplitudes to obtain classical probabilities. It is the procedure **R** and only **R**, that introduces uncertainties and probabilities into the quantum theory. (p. 250)

This passage clearly illustrates what I consider to be the author's gravest error of choice in his presentation of the material, namely, his use of self–chosen letter symbols (here, the letters **U** and **R**) to refer to, by way of

abbreviation, mathematical treatments which he considers to be too lengthy and/or technical to bother the reader with. In doing so, he automatically *guarantees* the impossibility of the reader attaining a level of comprehension of the material equivalent to his own, and also, thereby, makes all subsequent discussion less than fully comprehensible, as it must depend upon these obfuscating abbreviations. There are numerous examples of this thoughout the book. Thus, the reader, who brings with her/him a sincere desire to fully comprehend the author's intended meaning and to acquire a new level of understanding, is sure to be somewhat disappointted, and is also quite likely to become increasingly confused and frustrated. Despite the intrinsic fascination of the subject matter, this was, I regret to say, my experience in reading this book.

In the concluding section of the last chapter called "Where lies the physics of mind ?", Penrose writes:

> Some of the arguments that I have given in these chapter may seem tortuous and complicated. Some are admittedly speculative, whereas I believe that there is no escape from some of the others. Yet beneath all this technicality is the feeling that it is indeed 'obvious' that the conscious mind cannot work like a computer, though much of what is actually involved in mental activity might do so. (p. 448)

What is obvious to me is that, were some highly advanced 6th or 7th generation computer to be given the task of writing a book on Penrose's theme, it would probably be able to do so with greater logical lucidity and straightforwardness, but such a book might be lacking in an essential humanness that would make it readable. To me, however, this would be additional

support for Penrose's argument that human intelligence/mind/consciousness contains a mysterious quality that is essentially "uncomputable".

ADDITIONAL READINGS

For those who are interested in deepening their insight into quantum theory, the following books are recommended;

Feynman, Richard P. , *QED; The strange theory of light and matter,* Penguin Books, 1990.

Jauch, J. M., *Are Quanta Real ?:A Galilean Dialog,* Indiana University Press, 1989.

---

*The book's title alludes to Hans Christian Andersen's folktale "The Emperor's New Clothes".

*AI is abbreviation for "artificial intelligence".

*The Greek letter "psi" refers here to the "wavefunction" of a particle as specified by Shrödinger's equation.