Can Physics Provide a Theory of Consciousness?

A Review of *Shadows of the Mind* by Roger Penrose

Bernard J. Baars
The Wright Institute
2728 Durant Ave.
Berkeley, Calif. 94704
U.S.A.

baars@cogsci.berkeley.edu

PSYCHE, 2(8), May 1995
http://psyche.cs.monash.edu.au/v2/psyche-2-08-baars.html

KEYWORDS: awareness; cognitive theory; consciousness; Global Workspace theory; phenomenology; Quantum consciousness

REVIEW OF: Roger Penrose (1994) *Shadows of the Mind*. New York: Oxford University Press. 457 pp. Price: $25 hbk. ISBN 0-19-853978-9.

1. Introduction

1.1 Physics is surely the most beautiful of the sciences, and it is esthetically tempting to suppose that two of the great scientific mysteries we confront today, observer effects in quantum mechanics and conscious experience, are in fact the same. Roger Penrose is an admirable contributor to modern physics and mathematics, and his new book, *Shadows of the Mind* (SOTM) offers us some brilliant intellectual fireworks --- which for me at least, faded rapidly on further examination.

1.2 I felt disappointed for several reasons, but one obvious one: Is consciousness really a physics problem? Penrose writes,

A scientific world-view which does not profoundly come to terms with the problem of conscious minds can have no serious pretensions of completeness. Consciousness is part of our universe, so any *physical theory* which makes no proper place for it falls fundamentally short of providing a genuine description of the world. I would maintain that there is yet no physical, biological, or computational theory that comes very close to explaining our consciousness ... (emphasis added)

1.3 Having spent 17 years of my life trying to do precisely what Penrose suggests has not and cannot be done, this point was a bit disconcerting. But even more surprising was the

claim that consciousness is a problem in physics. The conscious beings we see around us are the products of billions of years of biological evolution. We interact with them --- with each other --- at a level that is best described as psychological. All of our evidence regarding consciousness depends upon reports of personal experiences, and observation of our own perception, memories, attention, imagery, and the like. The evidence therefore would seem to be exclusively psychobiological. We will come back to this question.

1.4 The argument in SOTM comes down to two theses and a statement of faith. The first thesis I will call the "Turing Impossibility Proof," and the second, the "Quantum Promissory Note". The statement of faith involves classical Platonism of the mathematical variety, founded in a sense of certainty and wonder at the amazing success of mathematical thought over the last 25 centuries, and the extraordinary ability of mathematical formalisms to yield deep insight into scientific questions (SOTM, p. 413). This view may be captured by Einstein's well-known saying that "the miraculous thing about the universe is that it is comprehensible." While I share Penrose's admiration for mathematics, I do not believe in the absolute nature of mathematical thought, which leads him to postulate a realm of special conscious insight requiring no empirical investigation to be understood.

1.5 After considering the argument of SOTM I will briefly sketch the current scientific alternative, the emerging psychobiology of consciousness (see Baars, 1988, 1994; Edelman, 1989; Newman and Baars, 1993; Schacter, 1990; Gazzaniga, 1994). Though the large body of current evidence can be stated in purely objective terms, I will strive to demonstrate the phenomena by appealling to the reader's personal experience, such as your consciousness of the words on *this page*, the inner speech that often goes with *the act of reading carefully*, and so on. Such demonstrations help to establish the fact that we are indeed talking about consciousness *as such*.

2. Has Science Failed To Understand Consciousness?

2.1 Central to SOTM is Penrose's contention that contemporary science has failed to understand consciousness. There is more than a little truth to that --- if we exclude the last decade --- but it is based on a great historical misunderstanding: It assumes that psychologists and biologists have *tried* to understand human experience with anything like the persistence and talent routinely devoted to memory, language, and perception. The plain fact is that we have not treated the issue seriously until very recently. It may be difficult for physicists to understand this --- current physics does not seem to be intimidated by anything --- but the subject of conscious experience, the great core question of traditional philosophy, has simply been taboo in psychology and biology for most of this century. I agree with John Searle that this is a scandalous fact, which should be a great source of embarassment to us in cognitive psychology and neuroscience. But no one familiar with the field could doubt it. As Crick and Koch (1992) have written, "For many years after James penned *The Principles of Psychology* (1890). most cognitive scientists ignored consciousness, as did almost all neuroscientists. The problem was felt

to be either purely "philosophical" or too elusive to study experimentally. In our opinion, such timidity is ridiculous."

2.2 Fortunately the era of avoidance is visibly fading. First-order theories *are* now available, and have not by any means been disproved (Baars,1983, 1988, and in press; Crick & Koch, 1992; Edelman, 1989; Gazzaniga, 1994; Schacter, 1990; Kinsbourne, 1993; etc.). In fact, there are significant commonalities among contemporary theories of consciousness, so that one could imagine a single, integrative hybrid theory with relative ease. But Penrose does not deal with this literature at all.

2.3 Has science failed, and do we need a scientific revolution? Given the fact that we have barely begun to apply normal science to the topic, Penrose's call for a scientific revolution seems premature at best. There is yet nothing to revolt against. Of course we should be ready to challenge our current assumptions. But it has not been established by any means that ordinary garden-variety conscious experience *cannot* be explained through a diligent pursuit of normal science.

3. A Critique Of The Turing "Impossibility Proof"

3.1 Impossibility arguments have a mixed record in science. On one side is the proof scribbled on the back of an envelope by physicists in the Manhattan Project, showing that the first enriched uranium explosion would not trigger a chain reaction destroying the planet. But notice that this was not a purely mathematical proof; it was a physical-chemical-mathematical reductio, with a very well-established, indispensible empirical basis. On the side of pure mathematics, we have such historical examples as Bishop Berkeley's disproof of Newton's use of infinitesimals in the calculus. Berkeley was mathematically right but the point was empirically irrelevant; physicists used the flawed calculus for two hundred years with great scientific success, until just before 1900 the paradox was resolved by the discovery of converging series.

3.2 Even more empirically irrelevant was Zeno's famous Paradox, which seemed to show that we cannot walk a whole step, since we must first cover half a step, then half of half a step, then half of the remaining distance, and the like, never reaching the whole intended step. Zeno of Elea used this clever argument to prove to the astonishment of the world that motion was impossible. But that did not paralyze commerce. Ships sailed, people walked, and camels trudged calmly on their way doing the formally impossible thing for a couple of thousand years until the formal solution emerged. And of course we have more than a century of mathematical reductios claiming that Darwinian evolution is impossible if you combine all the a priori probabilities of carbon chains evolving into DNA and ending up with thee and me. These reductios on behalf of divine Creation still appear with regularity, but the biological evidence is so strong that they are not even considered.

3.3 The problem is of course that a mathematical model is only as good as its assumptions, and those depend upon the quality of the evidence. The whole Turing Machine debate and its putative implications for consciousness is in my opinion a great distraction from the sober scientific job of gathering evidence and developing theory about the psychobiology of consciousness (e.g., Baars, 1988; 1994). The notion that the Turing argument actually tells us something scientifically useful is amazingly vulnerable. After all, the theory assumes an abstract automaton blessed with infinite time, infinite memory, and an environment that imposes no resource constraints. The brain is a massively parallel organ with 100 billion simultaneously active neurons, but the Turing Machine is at the extreme end of serial machines. This appears to be the reason why discussion of the Turing topic appears nowhere in the psychobiological literature. It seems primarily limited to philosophy and the general intellectual media.

3.4 Finally, it turns out that all current cognitive and neural models are formal Turing equivalents. That means the mathematical theory is useless in the critical task of choosing between models that are quite different computationally and on the evidence. It does not distinguish between neural nets and symbolic architectures for example, as radically different as they are in practice. But that is exactly the challenge we face today: choosing between theories based on their fit with the evidence. Here the theory of automata is no help at all.

3.5 A small but telling fact about Penrose's book caught my attention: of its more than 400 references, fewer than forty address the psychology or biology of consciousness. But all our evidence on the subject is psychological and, to a lesser extent, biological! It appears that Penrose's topic is not consciousness in the ordinary psychoneural sense, like waking up in the morning from a deep sleep or listening to music. How the positive proposals in SOTM relate to normal psychobiological consciousness is only addressed in terms of a technical hypothesis. Stuart Hameroff, an anesthesiologist at the University of Arizona currently working with Penrose, has proposed that general anesthetics interact with neurons via quantum level events in neural microtubules, which transport chemicals down axons and dendrites. It is an interesting idea, but it is by no means accepted, and there are many alternative hypotheses about anesthetics. But it is a real hypothesis: testable, relevant to the issue of consciousness, and directly aimed at the quantum level.

3.6 Penrose calls attention to the inability of Turing Machines to know when to stop a possibly nonterminating computation. This is a form of the Goedel Theorem, from which Penrose draws the following conclusion: "Human mathematicians are not using a knowably sound algorithm in order to ascertain mathematical truth." That is to say, if humans can propose a Halting Rule which turns out to be demonstrably correct, and if we take Turing Machines as models of mathematicians, then the ability of mathematicians to come up with Halting Rules shows that their mental processes are not Turing-computable.

3.7 I'm troubled by this argument, because all of the cognitive studies I know of human formal reasoning and logic show that humans will take any shortcut available to find a plausible answer for a formal problem; actually following out formalisms mentally is rare in practice, even among scientists and engineers. Human beings are not algorithmic creatures; they prefer by far to use heuristic, fly-by-the-seat-of-your-pants analogies to situations they know well. Even experts typically use heuristic shortcuts. Furthermore, the apparent reductio of Penrose's claim has a straightforward alternative explanation, namely that one of the premises is plain wrong. The implication psychologically is not that people are fancier than any Turing Machine, but that they are much sloppier that any explicit algorithm, and yet do quite well in many cases.

3.8 The fact that people *can* walk is an effective counter to Zeno's Paradox. The fact that people can talk in sentences was Chomsky's counter to stimulus-response theories of language. Now we know that people can in many cases find Halting Rules. It's not that human processes are noncomputible by a real computer --- numerous mental processes have been simulated with computers, including some formidable ones like playing competitive chess --- but rather that the formal straightjacket of Turing Machinery is simply the wrong model to apply. This is the fallacy in trying to attribute rigorous all-or-none logical reasoning to ordinary human beings, who are pragmatic, heuristic, cost-benefit gamblers when it comes to solving formal problems.

3.9 Penrose proceeds to deduce that consciousness is noncomputable by Turing standards. But even this claim is based only on intuition; the argument has the form, "mathematicians have an astonishingly good record gaining fundamental insights into a variety of formal systems; this is obviously impossible for a Turing automaton; hence mathematicians themselves cannot be modeled by such automatons." From a psychobiological point of view the success of mathematical intuition is more likely reflect the nervous system's excellent heuristics for discovering patterns in the world. The brain appears to have sophisticated knowledge of space, for example, which may in turn allow deep geometrical intuitions to occur with great accuracy in talented individuals. In effect, we may put a billion years of brain evolution of spatial processing to good use if we are fortunate enough to be mathematically talented.

4. The Quantum Promissory Note

4.1 Having proved that Turing machines cannot account for mathematical intuition, Penrose develops the idea that Quantum Mechanics will provide a solution. QM is the crown jewel of modern theoretical physics, an endless source of insight and speculation. It shows extraordinary observer paradoxes. Consciousness is a mysterious something human observers have, and many people leap to the inference that the two observer mysteries must be the same. But this is at best a leap of faith. It is much too facile: observations of quantum events are not made directly by human beings but by such devices as Geiger counters with no consciousness in any reasonable sense of the word.

Conscious experience so far as we know is limited to huge biological nervous systems, produced over a billion years of evolution.

4.2 There is no precedent for physicists deriving from QM any macrolevel phenomenon such as a chair or a flower or a wad of chewing gum, much less a nervous system with 100 billion neurons. Why then should we believe that one can derive psychobiological consciousness from QM? QM has not been shown to give any psychological answers. Conscious experience as we know it in humans has no resemblance to recording the collapse of a quantum wave packet. Let's not confuse the mysteries of QM with the question of the reader's perception of *this printed phrase* , or the inner sound of *these words* !

4.3 What can we make of Penrose's Quantum Promissory Note? All scientific programs are promissory notes, making projections about the future and betting on what we may possibly find. The Darwin program was a promissory note, the Human Genome project is, as are particle physics and consciousness research. How do you place your bets? Is there a track record? Is there any evidence?

5. Treating Consciousness As A Variable: The Evidence For Consciousness As Such

5.1 We are barely at the point of agreeing on the real scientific questions, and on the kind of theory that could address them. On the matter of evidence, Baars (1983, 1988, 1994 and in press), Libet (1985) and others have argued that empirical constraints bearing on consciousness involve a close comparison of very similar conscious and unconscious processes. As elsewhere in science, we can only study a phenomenon if we can treat it as a variable. Many scientific breakthroughs result from the realization that some previously assumed constant, like atmospheric pressure, frictionless movement,the uniformity of space, the velocity and mass of the Newtonian universe, and the like, were actually variables, and that is the aim here. In the case of consciousness we can conduct a contrastive analysis comparing waking to sleep, coma, and general anesthesia; subliminal to supraliminal perception, habituated vs. novel stimuli, attended vs. nonattended streams of information, recalled vs. nonrecalled memories, and the like. In all these cases there is evidence that the conscious and unconscious events are comparable in many respects, so that we can validly probe for the essential differences between otherwise similar conscious and unconscious events (See Greenwald, 1992; Weiskrantz, 1986; Schacter, 1990).

5.2 This "method of contrastive analysis" is much like the experimental method: We can examine closely comparable cases that differ only in respect to consciousness, so that consciousness becomes, in effect, a variable. However, instead of dealing with only one experimental data set, contrastive analysis involves entire categories of well-established phenomena, summarizing numerous experimental studies. In this way we can highlight

the variables that constrain consciousness over a very wide range of cases. The resulting robust pattern of evidence places major constraints on theory (Baars, 1988; in press).

6. Can Penrose Deal With Unconscious Information Processing?

6.1 Like many psychologists before 1900 Penrose appears to deny unconscious mental processes altogether. This is apparently because his real criterion is introspective access to the world of formal ideas. But introspection is impossible for unconscious events, and so the tendency for those who rely on introspection alone is to disbelieve the vast domain of unconscious processes.

6.2 Unconscious processing can be inferred from numerous sources of objective evidence. The simplest case is the great multitude of your memories that are currently unconscious. You can now recall this morning's breakfast --- but what happened to that memory before you brought it to mind? There is much evidence that even before recall the memory of breakfast was still represented in the nervous system, though not consciously. For example, we know that unconscious memories can influence other processes without ever coming to mind. If you had orange juice for breakfast today you may switch to milk tomorrow, even without bringing today's juice to mind. A compelling case can be made for unconscious representation of habituated stimuli, of memories before and after recall, automatic skills, implicit learning, the rules of syntax, unattended speech, presupposed knowledge, preconscious input processing, and many other phenomena. In recent years a growing body of neurophysiological evidence has provided convergent confirmation of these claims. Researchers still argue about some of the particulars, but it is widely agreed that given adequate evidence, unconscious processes may be inferred.

6.3 What is the critical difference then between comparable conscious and unconscious processes? There are several, but perhaps the most significant one is that conscious percepts and images can trigger *access* to unanticipated knowledge sources. It is as if the conscious event is broadcast to memory, skill control, decision-making functions, anomaly detectors, and the like, allowing us to match the input with related memories, use it as a cue for a skilled actions or decisions, and detect problems in the input. At a broad architectural level, conscious representations seem to provide access to multiple knowledge source in the nervous system, while unconscious ones seem to be relatively isolated. The same conclusion follows from other contrastive analyses. (See Baars, 1988).

6.4 None of this evidence appears to fit in the SOTM framework, because it has no role for unconscious but vitally important information processing. This is a major point on which the great weight of psychobiological evidence and SOTM are fundamentally at odds.

7. The Emerging Psychobiology Of Consciousness

7.1 The really daring idea in contemporary science is that consciousness may be understandable *without* miracles, just as Darwin's revolutionary idea was that biological variation could be understood as a purely natural phenomenon. We are beginning to see human conscious experience as a major biological adaptation, with multiple functions. It seems as if a conscious event becomes available throughout the brain to the neural mechanisms of memory, skill control, decision-makings, anomaly detection, and the like, allowing us to match our experiences with related memories, use them as a cue for skilled actions or decisions, and detect anomalies in them. By comparison, unconscious events seem to be relatively isolated. Thus consciousness is not just any kind of knowledge: It is knowledge that is widely distributed, that triggers off widespread unconscious processing, has multiple integrative and coordinating functions, aids in decision-making, problem-solving and action control, and provides information to a self-system.

8. Conclusion

8.1 I don't know if consciousness has some profound metaphysical relation to physics. Science is notoriously unpredictable over the long term, and there are tricky mind-body paradoxes that may ultimately demand a radical solution. But at this point in the vexed history of the problem there is little question about the preferable scientific approach. It is not to try to solve the mind-body problem first --- that effort has a poor track record --- or to pursue lovely but implausible speculations. It is simply to do good science using consciousness as a variable, and investigating its relations to other psychobiological variables.

References

Baars, B.J. (1983). Conscious contents provide the nervous system with coherent, global information. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation, 3,* 45-76. New York: Plenum Press.

Baars, B.J. (1988) *A cognitive theory of consciousness.* Cambridge, UK: Cambridge University Press.

Baars, B.J. (1994) A thoroughly empirical approach to consciousness. PSYCHE 1(6) [80 paragraphs] URL:http://psyche.cs.monash.edu.au/volume1/ psyche-94-1-6-contrastive-1-baars.html

Baars, B.J. (in press) *Consciousness regained: The new science of human experience.* Oxford, UK: Oxford University Press.

Crick, F.H.C. & Koch, C. (1992) The problem of consciousness, *Scientific American, 267(3),* 153-159.

Edelman, G. (1989) *The remembered present: A biological theory of consciousness.* NY: Basic Books.

Gazzaniga, M. (1994) *Cognitive neuroscience.* Cambridge, MA: MIT Press.

Greenwald, A. (1992). New Look 3, Unconscious cognition reclaimed. *American Psychologist, 47(6),* 766-779.

James, W. (1890/1983). *The principles of psychology.* Cambridge, MA: Harvard University Press.

Kinsbourne, M. (1993). Integrated field model of conscousness. In G. Marsh & M. J. Brock (Eds.), *CIBA symposium on experimental and theoretical studies of consciousness.* (pp. 51-60). London: Wiley Interscience.

Libet, B. (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8,* 529-66.

Newman, J., & Baars, B. J. (1993). A neural attentional model for access to consciousness: A Global Workspace perspective. *Concepts In Neuroscience, 2(3),* 3-25.

Penrose, R. (1994) *Shadows of the mind.* Oxford, UK: Oxford University Press.

Schacter, D. L. (1990). Toward a cognitive neuropsychology of awareness: Implicit knowledge and anosognosia. *Journal of Clinical and Experimental Neuropsychology, 12(1),* 155-178.

Weiskrantz, (1986) *Blindsight: A single case and its implications.* Oxford, UK: Clarendon Press.