Prof. A. J. Leggett of the University of Illinois at Urbana-Champaign is well known to condensed matter physicists for his contribution to the understanding of the low-temperature phases of ^{3}He, and his pioneering work on incorporating dissipation into the description of quantum phenomena such as tunneling and coherence. His current research interests include the nature of the order in high temperature superconductors – but in the midst of all this activity he has maintained an abiding interest in fundamental issues such as the nature of quantum measurement and the possibility that quantum theory itself might be incomplete. The following article is based on a lecture delivered at the Indian Institute of Science early this year, and represents his position on many basic issues which confront physics at its deepest level, including the limitations of the kind of explanations of phenomena usually offered in terms of fundamental constituents, quantum non locality, and possible tests for the applicability or otherwise of quantum theory at the macroscopic level. This extended and closely argued presentation of his viewpoint on so many basic interrelated issues should appeal to a wide spectrum of readers since it is not based on specialized technical arguments but on basic principles.

—Editor

As a Martian might see us: Subversive reflections on the practice of physics

A. J. Leggett

A few years ago, I was privileged to hear a lecture on the early history of quantum mechanics by one of the surviving founders of the subject. He started by showing a chronological table of some of the major milestones of the subject in the mid- and late twenties – de Broglie's thesis, the papers of Heisenberg and Schrödinger, the Solvay conferences, and so on. Then he said that to place his own experience in context he had marked two arrows on the table: the first, he said, indicated the point at which he had entered university, the second the point at which he had felt able to start to make independent contributions to the subject. At this there was an audible gasp from the audience: the arrows in question were a bare eighteen months apart.

Nowadays it is of course almost unthinkable that anyone, however talented, could start to make substantial contributions to any area of physics within eighteen months of first meeting the subject at university undergraduate level. For most of us, just to get to the stage when we can publish anything at all requires a long and tedious apprenticeship, and to become recognized as an expert even in a small sub-field can easily take half a lifetime. Perhaps this has always been the norm, and the period of violent intellectual upheaval in the mid-and late twenties, with its atmosphere of 'anything goes' was a rare historical anomaly. Be that as it may, the point I am leading up to is that for most of us practising physicists – and I suspect the same goes, to a greater or lesser degree, for the practitioners of any of the so-called 'hard' sciences, and perhaps of the soft ones too – by the time we are in a position to obtain a reasonable overview of even a small part of the subject, we have already invested so much of our lives in the discipline that we have a strong commitment, not just intellectual but in some sense emotional too, to the implicit rules and assumptions which govern its current practice; and any attempt to question or even to analyse these rules and assumptions is liable to provoke reactions ranging from a vague sense of impropriety to downright alarm and even outrage. On the other hand, those who make this kind of analysis their professional business – philosophers, historians and sociologists of science – may be well versed in the kind of knowledge one can get from textbooks about problems now regarded as solved, but only in rare cases do they have much of the personal experience of grappling with genuinely open questions which seems to me necessary to get a realistic perspective.

In spite of (or perhaps because of) these considerations, I believe it is important that we physicists should occasionally try to stand back and ask what exactly it is we are doing, and what kind of implicit assumptions we make in doing it. Here I am not thinking so much of the societal implications or impact of various branches of pure or applied physics, important as these are, but rather of the intellectual structure of the subject as such: What are our general prejudices about the behaviour of the natural world? What kind of question do we regard as askable? What kinds of answer are acceptable? What *a priori* constraints do we put on the experimental evidence we will accept? and so on. I believe it is important to ask these questions not just for the long term health of the subject itself, but also because, for many people outside

A J Leggett is in Department of Physics, Loomis Laboratory of Physics, 1110, West Green Street, Urbana, Illinois, 61801-3080, USA.

physics, the latter is the paradigm of a successful and quantitative scientific discipline, and in particular there has been a strong tendency among some practitioners of the social sciences to look on it as an exemplar which their own field should imitate. If indeed they wish to do this – and I express no view on the desirability or otherwise of such a program – we as physicists at least have a duty to see that they know just what it is they are imitating!

For this purpose it is useful to introduce the idea of 'Martian'. A Martian, in the sense in which I shall use the term, is, like a perfect heat bath, an intellectual construct designed to permit a thought experiment: he/she/it* is a being of such superior intellectual ability that he can grasp all the subtleties of the latest advances in (terrestrial) physics without having to put in the long years of apprenticeship which we humans must undergo, and would therefore be under no temptation to accept things as natural simply because they have been part of his environment over all these years. Also, such a being, not being himself constrained by human biology or psychology, might be specially aware of the way in which these factors may affect our development of the apparently impersonal subject of physics. Let's try to imagine that such a being has thoroughly familiarized himself with the contents of Physical Review and other journals over the last few years, and has thoroughly understood all the technical details. Which features of our current practice might strike him as remarkable?

A couple of general remarks before I start. First, I have used in my title the word 'subversive', and indeed the examination of assumptions which are normally taken for granted must, I think, always have a certain subversive element in it, if only because very often the mere act of formulating an assumption explicitly has the effect of making it seem less self-evident than it appeared when cloaked in ambiguity. But the fact that one recognizes one's working assumptions as less than self-evident need not imply that one has a ready alternative to them. Indeed, it could be that our fictitious Martian would conclude that we *have* to do physics the way we do because of the kind of beings we are. Even if it eventually turns out to be possible to do it some different way, the saying attributed to Louis Armstrong is apposite: When asked by a fan 'where jazz was going?' he allegedly replied, 'Man, if I knew where jazz was going, I'd be there already!' Obviously, if I or anyone else had any concrete ideas on how to transcend the assumptions which our Martian might find questionable, we'd be doing it, not talking about it – and in a few very specific cases, we already are. As for the rest, the conscious and explicit recognition of the framework within which we currently operate may be an essential first step in this direction.

Secondly, while in this talk I will repeatedly question, at least by implication, the idea that there is a single 'correct' description of the physical world, and that current physics is engaged in progressively revealing it, I would be extremely unhappy if I were to be understood as embracing or endorsing a view which seems to be fashionable among some current practitioners of the sociology of science, namely that in the last resort scientific 'truth' and 'progress' is nothing more than a matter of social and intellectual convention. Perhaps in my first few years as a theoretical physicist, when like many people at such a stage I was essentially doing (or at least felt I was doing) applied mathematics with little real empirical content, I might have credited this thesis with some marginal plausibility; today, after many years of having conjectures about the real world both confirmed and refuted in real life experiments, it seems to me, as I suspect it does to most practising scientists, so silly as not to be worth serious discussion.

Let's start with an assumption which I suspect is made subconsciously not just by physicists but by just about all practising scientists, at least in the physical sciences†, and indeed to a very large extent by the man in the street going about his daily business; namely that there exists a fairly rigid distinction between on the one hand the 'hard' data which are the subject matter of sciences such as physics, and which can be verified by any competent and responsible observer, and on the other, those phenomena which are not so inter subjectively verifiable: the latter we class as illusions, hallucinations, fantasies etc. Is this distinction in fact as rigid as we like to think? To make this point picturesquely, let me imagine that I were to ask readers of this piece whether or not they believe in ghosts. I suspect that I would get rather few affirmative answers: The majority of readers, I imagine, do not believe in ghosts. But what, exactly, is it in which they are declining to believe? Here is a tentative definition of a ghost: a phenomenon whose observation requires *both* the presence of certain objective physical conditions (in the crudest cases, the right haunted house, etc) *and* of a certain mental disposition on the part of the observer. Most people, and certainly most physicists, would I suspect, agree to the proposition that ghosts, as so defined, do not exist: by which I assume they would mean that whenever such phenomena are reported, they are in fact a function exclusively of the mental state of the observer and have no objective physical correlates (apart, perhaps, from trivial and uninteresting ones such as the atmospheric conditions required to give rise to the 'hallucination' in question, etc). Would our Martian observer take this in his stride, I wonder? Or would he be sufficiently impressed by, for example, tales of alleged anomalous communication between identical

*From now on I for definiteness refer to the Martian as 'he', but this is, as the lawyers say, without prejudice to the eventual determination of the gender of the being in question.

†Perhaps some neuropsychologists might take a different view.

twins to find it somewhat remarkable that we are so confident in our present understanding of the relationship between what we conventionally class as the 'mental' and 'physical' worlds? Let me emphasize that I am not expressing any particular opinion about the degree of plausibility attaching to these or any other accounts of allegedly 'paranormal' phenomena; I am merely pointing out that these are certain kinds of hypothesis, which our Martian might not find *a priori* totally ridiculous, which we nevertheless – most of us – tend simply to exclude *a priori* without argument, simply because they do not fit neatly into the framework of science as we are accustomed to practise it.

Let's go on to some specific areas of physics. Most of us (not merely its practitioners!) are accustomed to regard high energy physics as the most 'fundamental' area of the subject, and there are innumerable locutions which reinforce this perception: high-energy physics is said to 'probe the ultimate constituents of the universe', 'unlock the most intimate secrets of Nature', 'peel the layers off the cosmic onion' and so on. Why so? Needless to say, it's not high energies as such that we regard as fundamental; rather, we are relying implicitly on the following argument. In wave optics, to 'see' something on a scale of order a we need a wave whose wavelength λ is itself of order a or smaller. But according to de Broglie a particle of momentum p has associated with it a wave of wavelength $\lambda = h/p$; and moreover, according to Einstein, in the relativistic limit momentum p implies energy $E = cp$. Hence, to resolve anything on a scale a we need to fire at it particles of energy of order hc/a: the smaller the scale, the larger the energy required. Of course there are other more technical considerations too, associated with eg. the instability of heavy particles; but in the last resort, when we assert, as most particle physicists would, that low-energy behaviour is a consequence of high-energy behaviour (which is therefore the more 'fundamental' area of research) what we are really claiming is that behaviour on the large scale is a consequence of behaviour on the smaller scale; or, to put it more bluntly, that the behaviour of big things follows from the behaviour of the smaller things of which they are composed. We have analysed the properties of macroscopic bodies – so goes the conventional wisdom – in terms of those of atoms and molecules: these system in turn behave as they do because of the properties of the electrons and nuclei: the behaviour of nuclei is determined by that of their constituent nucleons: and now we trace the properties of the nucleon itself to that of its constituent quarks. What could be more obvious than that the behaviour at each level is determined by that of the constituents at the next level below? and that therefore that area of physics which deals with the smallest constituents currently known, is the most fundamental?

Would our Martian indeed regard it as obvious, I wonder? Or would he, perhaps, reflect that for most of us – at least for most of us who have grown up with the technology of the late twentieth century – our experience of 'understanding how things work' starts with mechanical devices made by other human beings, and that the most natural way of achieving such an understanding is precisely to take the device apart into its constituent parts, since these are what the maker started with? Does this experience subconsciously colour our perception of what constitutes an 'explanation' of natural phenomena as well as of human artefacts? I will argue later by implication that this is indeed the situation in at least one important area of physics.

Be that as it may, let's try to stand back and abstract from all the technical details what it is that contemporary particle physics is actually doing. First, on the experimental front: apart from a few brave souls who are engaged in 'table-top' searches for axions, monopoles, massive neutrinos and the like, most high-energy experimentalists are engaged in a single enterprise which, conceptually if not technically, has a very simple structure. Namely, they accelerate particle A and particle B so as to hit one another, and watch where they and/or particles C, D, E ... emerge, and with what energy and (sometimes) spin. In other words, they measure a set of differential cross-sections. Now notice that this whole operation is quite deliberately and consciously set up so that all the interesting things happen only in the small region where the particles collide; the interpretation of the experiment is insensitive, by design, to the details of the working of the accelerators and detectors involved. In particular, the experiment is designed so that, as nearly as possible, the incoming beams are each described by quantum-mechanical pure states of definite momentum; and while the theory certainly predicts that, in certain cases at least, the outgoing states are not simple classical 'mixtures' of products of plane wave states, but have built into them subtle quantum correlations of the type which are important in Bell's theorem, the whole setup is designed precisely so that such subtleties can be neglected. No doubt this is sensible and indeed at the present stage probably necessary. But when our Martian observer hears the often-repeated claim of particle physicists that experiment has shown that Nature is actually simpler at high energies than we have any right to expect, might he not wonder whether this is due, at least in part, to the fact that we have chosen to ask her only questions which by their very construction, allow no subtlety in the answers?

Turning now to the theoretical front, what is it that high-energy theorists are actually doing? As always, there are exceptions to my description: but the main line of research in contemporary high-energy theory could, I think, reasonably be characterized as follows. One accepts the general framework of quantum field theory, usually though not always in the Lagrangian formulation, and one then tries to postulate a set of fields, and a Lagrangian which is a function of them, such that: (a) the Lagrangian respects the symmetries we believe to hold in nature: (b) the theory is mathematically self-consistent (no ultraviolet divergences etc) (c) you can actually calculate some of its implications for observable quantities such as cross-sections etc, and, ideally (d) the predicted values agree with existing experiment. Increasingly over the past twenty years or so, the quest for mathematical self-consistency has loomed large in this program, the most spectacular example probably being the way in which spontaneous symmetry breaking was shown to be an essential element of the electroweak unification scheme. A few years ago, at least, there were high hopes (I am not clear how far those at the forefront of the field now share them) that in the 'super-string' picture the constraints imposed by the need for self-consistency would be so severe that they would uniquely determine the parameters of the theory, including as outputs not only the masses and coupling constants of the known elementary particles but even the 'true' dimensionality of space-time. If such a program should be successful, then, it is said, we shall have a 'Theory of Everything': the Universe *has* to be exactly as it is.

Leaving aside the question of the likely success or otherwise of this enterprise, I can imagine our Martian friend becoming distinctly puzzled at this point. 'As I see it (he might say) the situation is this: You have a formalism – let us say that of Lagrangian field theory in its 'superstring' implementation – which you believe should in some sense correspond to the physical world. And you are trying to show that there is only one version of such a theory which is mathematically self-consistent, and that this version predicts the occurence of exactly the world we know. Okay, suppose you succeed in this program: so what? You would then say 'the true dimensionality of space-time is (let us say) 26, *because* in no other dimension do the equations of my chosen form of quantum field theory avoid divergences'. What kind of 'because' is that? Can mathematics – a subject which is usually taken to be concerned with analytic truth – really put constraints on how Nature can behave? Why should she not select (say) a 37-dimensional space-time, and simply refuse to be described by your form (or indeed perhaps any form) of quantum field theory at all?' Now I think that most of us would feel that our Martian colleague had in some sense missed the point; but to clarify what that point is we might have to analyse exactly what it is that we mean by an 'explanation' or 'theory' in physics, and I'll return to that question in a moment.

Before leaving high-energy physics, however, let's just look briefly at one more orthodoxy. A few years ago, a popular article in *Scientific American* on high-energy physics, after reviewing our current knowledge of the strong, electromagnetic, weak and gravitational interactions, assured its readers that 'the four basic interactions... together with cosmology, account for all known natural phenomena'. Note the bare form of the verb: not 'may account for' or even 'are thought to account for' but simply 'account for'. Now taken literally that statement is of course self-evidently and trivially false: there are a myriad of 'known natural phenomena' ranging from ball lightning to the phenomenon of human consciousness, which by no possible stretch of the imagination could be said to be at present 'accounted for' in terms of the four fundamental interactions – even in the diluted sense in which, say, many of the phenomena of solid state physics *are* so accounted for. (I'll return to the meaning of that statement later). How then shall we interpret claims like the one quoted? It is, of course, not really a statement of fact at all but an act of faith. What we are saying, in effect, is that at present there exists no natural phenomenon which we (a) believe to be scientifically reputable and (b) can prove cannot be explained in terms of the four fundamental interactions; and that in this situation the methodological principle known as Occam's razor dictates that we should make do with these four unless and until some such phenomenon turns up – which we believe (act of faith!) will not happen.

At which point, I suspect, our hypothetical Martian observer will promptly accuse us of circularity. Is it not true, he will enquire, that there are certain phenomena which you exclude from the category of 'scientifically reputable' for no better reason than precisely because you believe that they cannot be accounted for in terms of the known interactions? Take for example the phenomenon vulgarly known as 'spoon-bending' (or to its advocates as 'anomalous metal softening'). Granted (he will say) that there is any amount of showmanship, self-deception and perhaps in some cases downright fraud associated with this business – just as there was, historically speaking, with the medieval alchemy which we now recognize as the ancestor of the modern and eminently reputable science of chemistry – is it really on those grounds alone that most physicists dismiss it? Is it not also – and perhaps mainly – because they have convinced themselves, rightly or wrongly*, that there are good general reasons why such phenomena could not be accounted for in terms of any of the known laws of electromagnetism, gravitation, etc.? Let's suppose for the sake of argument – he might say – that the so-called 'fifth force' postulated a few years ago had turned out to be real. Now I can see that you humans would have believed in its reality only on the basis of experiments in particle physics or possibly solid-state physics or some other well-established branch of the subject.

*I personally suspect wrongly, but that is another story

But having once believed, I would bet a large amount that you would thereafter have rapidly 'discovered' (or re-discovered!) a whole series of phenomena – quite possibly including some you now class as 'paranormal' – which have in the past been dismissed as the product of fraud, hallucination or experimental error, but which are precisely accounted for by the new force. (At the time when the reality of the fifth force seemed an open question, I would have happily gone along with our Martian colleague in that bet).

Let's move on to another 'frontier' area of physics – cosmology. Here, we see a great deal of effort going into investigation of the question 'Why is the Universe the way it is? why do the galaxies have the particular shapes and distributions they do? why is the number of baryons the way it is, relative to the number of photons?' etc., etc. Now, five hundred years ago in medieval Europe, thinkers under the influence of Aristotle would have answered such questions (had they had the experimental information to ask them) by attempting to demonstrate that the arrangement we see is the 'best', or most perfect, possible one. Today, by contrast, most main-line cosmologists would regard the question as at least partially answered if they can show that a plausible set of initial conditions on the Universe, plus extrapolation of the laws of physics as we know them here on earth to conditions different by many orders of magnitude in temperature, density, etc., indeed leads to the Universe we see around us today.

Again I can imagine our extraterrestrial colleague getting restive. Not only might he marvel at our temerity in assuming that (say) a rise in temperature by a factor of 10^{20} should not introduce qualitatively new laws of physics (having marvelled, he might reflect that, if we want to get anywhere at all in cosmology, we probably have to make this kind of assumption in the absence of experimental evidence to the contrary), but, more seriously, he might challenge the concept of 'explanation' which is implicit here. 'Wait' he might say, 'all your textbooks tell me that the laws of physics, as you understand them, are at a fundamental level symmetric with respect to the arrow of time. The motion of a dynamical system, be it classical or quantum, is as completely specified by giving the final conditions as the initial ones. Why, then, do you claim to 'explain' the present state of the Universe by reference to its past history? Why not, rather, 'explain' the past by reference to the present or even the future? Are you not again indulging in a subtly anthropomorphic prejudice here, in that part of your basic, pre-scientific concept of 'explanation' is related to humanly activated processes where it is the initial conditions, and those alone, which are set by human agency?' In reply we might perhaps make two points. The first is an obvious practical one: while we may make all sorts of theoretical speculations, correct or otherwise, about the future of the Universe, our experimental information relates exclusively to its present and past; and it therefore makes some sense to ask whether all the information we have about its past history and present state can be made consistent within a given theoretical framework, whereas a similar question about the future cannot meaningfully be asked. The second point is that within our current conceptions about irreversibility there is indeed some logic in explaining the present with reference to the past rather than vice versa, in that the relationship is asymmetrical: if we use the thermodynamic concept of 'state', then many different initial states can lead to the same final state, but not vice versa, so that while it makes sense to ask 'which of the various initial states of the Universe which could have led to the present one was actually realized?' the converse of this question makes no sense*. Of course, our Martian might then challenge us as to why we believe what we do about irreversibility and the arrow of time, but that question is of a rather different nature.

At this point let us step aside for a moment and reflect on what it is we are trying to do when we construct a 'theory' in physics, and what we regard as the criterion for success. Of course, so long as we are working within a given conceptual framework – roughly speaking, what T. S. Kuhn calls a 'paradigm' – such as Lagrangian quantum field theory, there is no great problem: our job is simply to find the ingredients which make the theory self-consistent, tractable and in accordance with experiment, and while it goes without saying that this task is highly nontrivial, at least we have fairly well-defined and generally agreed criteria for what constitutes success. But let's ask the question at a broader level: What kind of relationship do we look for between the physical world and our theories of it? Perhaps if we can answer this question, we will be a little closer to understanding just what force some of our Martian colleague's doubts and reservations have.

Needless to say, the question of the relationship between the external world and the scientist's description of it is one which has exercised generations of philosophers of science – it is, perhaps, *the* central question of the discipline – and I have nothing particularly profound or original to contribute to it; I would like merely to present an analogy which, while possibly philosophically naive, may nevertheless, I believe, capture the implicit understanding of many working scientists as to what it is that they are trying to do. What I would like to suggest is that we should think of scientific theories as analogous to *maps* of the physical world (notice that I say 'maps', plural, not 'a map'.) What is a map? In the first place, it in not a picture of anything. Rather it is a representation in more or less symbolic form of certain interesting features of a particular

I owe this observation to Dr J. P. Sethna geographical region. Think of the various kinds of maps we know: there are Geological Survey maps, real estate agents' maps, road maps put out by the motoring organizations, military maps, maps of the Paris Metro system or of Chicago's O'Hare airport ... and so on. If we really think about it, they have rather few specific features in common, apart from being two-dimensional representations on pieces of paper. What they do have in common is that each conveys, in the form of a visual Gestalt, that information about the region in question which is relevant for the purpose for which it is designed. Which features of the external physical world are represented in the map depends on its intended function; it need not necessarily preserve metric or even topological relationships†. Mountaineers do not usually complain because their maps fail to show the names of roads, nor travellers on the London Underground because, if you believe the wall maps, all lines run in one of eight discrete directions: each map is perfectly adequate for the purpose for which it is designed. There is not and could not be a 'perfect map'!

One thing about the analogy between maps and scientific theories needs to be specially emphasized: A map is a representation which is constructed *by human beings* with specific human purposes in mind. In so far as the analogy is a good one, therefore, many of the features of our present-day science which puzzle our Martian friend may perhaps be traced not to any 'intrinsic' or 'objective' properties of the world, but rather to the constraints imposed on what for us human beings is an adequate description of it by the capacities and limitations of our own minds. Needless to say, this idea is not exactly a novel one – it goes back at least to Immanuel Kant – but it seems to me that it may be desirable to reemphasize it in the intellectual context of late twentieth-century physics. Perhaps it may, inter alia, shed some light on why we think of 'simplicity' as such a virtue in a scientific theory, and why the 'unification' of apparently disparate natural phenomena has always been such a major goal in physics: it is not so much that there is any *a priori* reason why Nature should be simple or unified, but rather that it saves us an enormous amount of mental filing space! A 'good' and simple scientific theory, like a good and simple map, gives us the ability to get where we want to, physically, intellectually or technologically, quickly, cleanly and without unnecessary diversions – and that, surely, is a large part of what we mean by 'understanding' a phenomenon in physics. One obvious implication of this point of view is that it may often be more valuable to have a simple theory which gets things approximately right than a complicated one which gets them exactly right (just as no motorist in his right mind would use a series of 6" U.S. Geological Survey maps to get from New York to Los Angeles!) Another intriguing question which it suggests is whether we shall in the end ever really be satisfied with a theory – such as QCD in its current state of development – in which, while the basic postulates can be simply and clearly stated, it is largely impossible to obtain results which can be compared with experiment without recourse to large-scale numerical computation. My guess is that in the end we shall come to regard such theories as second-best, in that they do not give us any real 'understanding' in the sense in which we have been used to it in physics; but this may no doubt depend on the extent to which we eventually get used to the idea of using computers as a genuinely symbiotic extension of our own mental powers, something which seems to me difficult to predict.

Let's go back for a moment, in the light of our map-making analogy, to our Martian friend's skepticism about quantum field theory. Let's again suppose that it is conclusively shown that some version of this theory – perhaps the 'superstring' version, perhaps something different – is mathematically self-consistent and anomaly-free only, say, in 26 dimensions. What would this show? From the present point of view, we have been trying to make a particular kind of map of the physical world using certain particular tools. It is not the only possible kind of map, and the reasons we are trying to make this particular kind are mainly historical; moreover, we are in no way guaranteed *a priori* that it will be possible at all. If we can show that the map is a successful one, then we can infer that the world is indeed 'really' 26-dimensional, and that of course is an interesting and nontrivial piece of information about it. But to say that we have in any sense *explained* the 26-dimensionality of space-time in this way seems to me to put the cart before the horse – what we have actually discovered, mostly, is an interesting fact about the internal structure of our chosen form of quantum field theory. Here, as elsewhere, we physicists may be in danger of unconscious narcissism – we marvel at the beauty we are revealing in the Universe, oblivious that a large part of that beauty resides in the probably transient mathematical structures which are our current best attempt to reflect its topography.

The considerations I have advanced above become, I believe, particularly important when we come to the last area of modern physics on which we shall let our Martian colleague loose, namely the physics of condensed matter. Here, after reading a few years' worth of *Physical Review B* and similar journals, I suspect he would be very puzzled indeed; for he would almost certainly have come to the conclusion that what we are actually doing

*Though I suspect that future generations may use three-dimensional holographic maps for some purposes.

†If you are a traveller departing from Charles de Gaulle airport in Paris, it will not matter to you if your airport map fails to represent the complicated topology of the connection tubes, provided it gets you to your satellite. (If on the other hand you are an aiport security officer, the topology is vital.)

in condensed-state physics – at least in the theoretical side of the subject – is not at all what most of us in the field, and just about everyone outside the field, *thinks* we are doing. Roughly speaking, the conventional picture might go as follows: Suppose we want to explain a particular phenomenon observed experimentally in bulk condensed matter – say the magnetic properties, or superconductivity, or the viscosity of a particular kind of liquid as it is cooled into the glassy state. What we do is: We first get our experimental colleagues to determine the exact microscopic composition of the system in question in terms of its constituent atoms, etc. Next, we write down a 'microscopic' description of the behaviour of the system in terms of these constituents – for example, we formulate a Hamiltonian in terms of the atoms and electrons. So, in principle, we have a grand many-body Schrödinger's equation for the 10^{23}-odd particles composing the system. We then sit down and try to solve this equation so as to predict values for macroscopic properties such as magnetization, conductivity, etc. which we can compare with experiment. Of course, because of the extreme complexity of the equation we cannot hope to get an exact solution to it; so we are forced, regrettably but inevitably, to make approximations to the solution. Fortunately, however, even if our analytical problem-solving abilities do not improve much with time, the computing power available to us nowadays does; so we can expect to make steady progress towards the ultimate goal of completely explaining the behaviour of complex, macroscopic bodies in terms of their microscopic constituents.

At first sight this description looks so innocuous as to be totally uncontroversial; and certainly it is one which is implicit in the *words* that most condensed state theorists write in the abstract and conclusions to their papers. Yet I think that our Martian might find it far from innocuous. In the first place, he might point out, from your description you are working towards a goal which, should you ever achieve it, would be utterly and totally useless. Let's just suppose, for the sake of argument, that you could ever know in complete detail the exact microscopic constitution of your system in terms of individual atoms, etc. (Remember, this would require you to know the exact shape of the surface, the position of every single chemical impurity, dislocation, etc. and so on – but let that pass). And suppose moreover that available computing power increases so spectacularly that you can actually solve numerically, to a sufficient degree of accuracy, the Schrödinger equation for your 10^{23}-particle system. What conceivable good would that do you? You will be faced with a pile of millions of tons of computer printouts – or, optimistically, computer graphics – which will be totally useless to you without some consciously chosen principle of organization. It is conceivable, if unlikely, that this operation might have some point in the context of purely technological applications, but surely it cannot have anything to do with what you mean by *understanding* the behaviour of a complex, macroscopic system.

Secondly, he might say, let's take a look at those so-called 'approximations' you keep making. In mathematics the word 'approximation' has a clear and well-defined sense: one has, perhaps, some small parameter and shows, by the rigorous process of deductive reasoning characteristic of mathematics, that replacing a particular function by the first few terms in its Taylor expansion incurs an error whose maximum value is some positive power of that small parameter and tends to zero with it. Now compare a typical 'approximation' of condensed-matter physics, say the random-phase approximation. It is not an approximation in the mathematical sense at all; rather, it is an *educated guess* that the deletion of certain terms in the equations of motion will not radically affect the solution. No small parameter is involved here (at least not necessarily so), nor any process of deduction which would be recognizable as such by a professional mathematician. So the idea that you obtain the macroscopic behaviour from microscopic equations by a process of 'approximation' is actually quite misleading; what you are in fact doing, at best, is to show that there are reasonable guesses which you can make about the solutions to the microscopic equations which are compatible with what you know to be the experimental macroscopic behaviour.

Thirdly, he might say, even with these provisos I challenge you to show me one single case when a major advance in understanding in condensed-matter physics has come about as a result of derivation from a first-principle microscopic description. Now this is likely to raise cries of protest from those knowledgeable about the area: isn't the BCS theory of superconductivity, to take just one case, a spectacular example of just such an advance? Yes and no. Certainly, BCS were able to visualize and interpret the macroscopic phenomenon of superconductivity in terms of a picture of what the microscopic constituents – the electrons – are doing. But they precisely did *not* derive their results from a complete microscopic description of the 10^{23}-odd atoms and electrons composing the superconductor. Rather, a very large part of their achievement was to isolate, from a vast and undifferentiated mass of information, precisely those special features – such as the existence of an attraction between electrons of opposite momentum and spin – which were relevant to the phenomenon in question, and to use it to build a model which, while certainly not adequate for all purposes*, reflects all the considerations essential in the context of superconductivity. This state of affairs seems to me very typical. To be sure, there are plenty of papers in the 'physics'

*No-one in his/her right mind would use the BCS model to predict X-ray scattering from a metal literature which do nothing but mathematics, and in a few cases even rigorous mathematics – but all the physics has then gone into the choice of the original model.

In this situation I believe that it is sensible to reorient our view of the kinds of questions that we are really asking in condensed-matter physics. Rather than chasing after the almost certainly chimerical goal of *deducing* the behaviour of macroscopic bodies rigorously from postulates regarding the microscopic level, it may be better to view the main point of the discipline as, first, the building of autonomous concepts or models at various levels, ranging all the way from the level of atomic and subatomic physics to that of thermodynamics; and, second, the demonstration that the relation between these models at various levels is one not of deducibility but of *consistency* – that is, that there are indeed 'physical approximations' we can make which make the models at various levels mutually compatible. From this point of view, indeed, the most important theorems in physics are theorems about the *in*compatibility of models at different levels; one such is the well-known theorem of Bohr and van Leeuwen, which states that no model of atoms which employs only classical mechanics and classical statistical mechanics can produce the observed atomic diamagnetism; and an even more spectacular example is Bell's theorem.

Let's finally go back to our 'map-making analogy'. Suppose one wanted to represent in compact form the transportation network of a country like England. There are at least two ways one could go about it. On the one hand, one could take an extensive series of aerial photographs and reduce them in scale. In this way one would in some sense have started with an 'exact' picture and made 'approximations' to it; the degree to which objects would be accurately represented in it would be a function simply of their physical size, and would in no way reflect their intrinsic importance in the communications network. Such a picture would be of little use as a practical guide. A much more useful alternative would be simply to draw a road or rail map of the type with which we are all familiar – that is, something which in no sense pretends to be a picture, but rather is a schematic representation of that information which is important in the context. Naturally, the map must be consistent with the topography shown in the detailed aerial photographs, but it is in no sense an *approximation* to them; indeed, it may embody quite different kinds of information (for example, British road maps conventionally represent roads by different colours in accordance with the Ministry of Transport classification, something which no amount of inspection of aerial photographs could reveal!). The view which I have been arguing against (which seems to be widespread among philosophers of science and other onlookers, and more surprisingly even among some practising scientists) in effect holds that the theories and concepts of condensed-matter physics are analogous to the reduced aerial photographs; my claim is that they are much better viewed as analogous to the map.

Well, okay, we might say to our (irritatingly persistent) Martian interlocutor. Given that it is not necessarily the goal of condensed-matter physics to deduce the properties of macroscopic bodies from those of their microscopic constituents, it is surely nonetheless true that the former are a consequence of the latter? If in fact we were able to solve Schrödinger's equation for the 10^{23} odd atoms which constitute a piece of iron or a small biological organism, surely it would at least give a correct – if not necessarily very useful – description of its macroscopic behaviour? That is, there are no new laws of physics which we need to describe the behaviour of complex, macroscopic bodies, over and above those which describe their atomic-level constituents. Surely our Martian friend would at least have to agree with *that*? After all, here are a couple of the most distinguished particle physicists in the world on this subject: 'Important theories do emerge in other sciences [than high-energy physics and cosmology] ... How truly fundamental are they? Do they not result from a complex interplay among many atoms, about which Heisenberg and his friends taught us all we need to know long ago?' (S. Glashow, *Physics Today*, 1986, Feb., p. 11).

'No-one thinks that the phenomena of phase transitions and chaos ... could have been understood on the basis of atomic physics without creative new scientific ideas, but does any one doubt that real materials exhibit these phenomena because of the properties of the particles of which the materials are composed?' (S. Weinberg, *Nature*, 1987, **330**, 435). Now the interesting thing about these two quotations is that both use the device of the rhetorical question. And whenever I hear a rhetorical question, whether in physics, politics or elsewhere, I am reminded of the (possibly apocryphal) story of the UN diplomat who takes home the script of a colleague's speech by mistake. On leafing through it he is surprised to find that it is annotated like a piece of music – 'crescendo', 'andante', etc; but he is particularly amused to come across a point at which the author has annotated the margin 'weak point. SHOUT!' Would our Martian friend just nod at the above quotations, I wonder? Or would he wonder whether it really is as obvious as Glashow and Weinberg take it to be, that the behaviour of complex bodies is entirely determined by that of their constituents?

This is probably the hardest point in this whole article to get across. In the first place, reductionism is probably as deeply ingrained in the thinking of most of us as any single element in the whole of our scientific world view. Secondly, most people are rightly impressed by the extent to which microscopic ideas, and particularly those associated with quantum mechanics, have given us insight, not just into much of chemistry but also into macroscopic phenomena which from a purely macroscopic viewpoint are quite mysterious, such as ferromagnetism and superconductivity. In the light of this record of success, is it not perverse to the point of eccentricity to cast even a shadow of doubt on the reductionistic hypothesis?

Let me start by emphasizing one point: So long as one is dealing with those phenomena, and only those, where we believe that the predictions of quantum mechanics are well approximated by those of classical physics, then I believe that the evidence for the reductionist point of view is very strong, and moreover there is absolutely no *a priori*, internal reason to challenge it. It is only when we come to intrinsically quantum phenomena that we have a problem. First, I would claim, the positive evidence in favour of reductionism in this regime is much less strong than it looks at first sight, and secondly – and much more important – there are indications which are intrinsic to the quantum formalism itself that the reductionistic program not only might but must eventually fail. Clearly, both these claims need amplification.

As regards the positive evidence, I believe that if we think about it carefully we will come to the conclusion that in just about every case in which we are observing the effects on the macroscopic scale of intrinsically quantum-mechanical behavior of the constituent atoms and electrons, be it in chemical reactions, magnetism, superconductivity, laser action or whatever, the observed macroscopic effect is in some sense the sum of one – or at best few-particle behaviour at the atomic level. For example, in a typical 'macroscopic quantum effect' in the conventional sense, such as the Josephson effect, what we are actually seeing is the effect of a macroscopically large number of Cooper pairs behaving in identical fashion; the observed supercurrent is simply the sum of the supercurrents carried by the individual pairs of electrons. Similarly, in laser diffraction, we are simply seeing the coherent sum of the behaviour of many individual photons. (To be sure, there exist phenomena in condensed-matter physics which cannot be explained as simply the sum of single particle or few-particle behaviour, perhaps the most obvious example being the thermodynamic behaviour near a second order phase transition: but it is significant that this behaviour is generally believed to be quite insensitive to quantum effects, except perhaps in rather trivial ways*.) So long as we are dealing with the summed effects – even the summed quantum effects – of a large number of small groups, there seems no reason to doubt that nature tolerates a reductionist approach; but this, if one thinks about it, is hardly surprising. To conclude from this that *all* phenomena in complex bodies have a reductionist explanation seems to me rather as if a sociologist should establish to his/her satisfaction that the total food consumption of a nation is equal to the sum of the consumption of the individuals comprising it, and then go on to infer that properties such as political or religious behavior must likewise be explainable entirely in terms of the properties of isolated individuals. Could it be – our Martian might ask – that the very nature of the experiments we conventionally conduct on macroscopic systems subtly biases the result in favour of a reductionist interpretation?

Be that as it may, the principle of Occam's razor certainly favours reductionism, in physics if not in sociology, and it would indeed be perverse to maintain the opposite point of view were there not some positive reason to do so. I now therefore turn to my second point – that the very structure of the quantum-mechanical formalism which we (most of us!) regard as describing the physical world itself gives strong positive hints in this direction.

Let's us start with a phenomenon which actually turns out to be fairly innocuous in the present context, but is nevertheless surprising and a hint of what is to come – the phenomenon usually known as the Aharonov–Bohm effect. In this, the current flowing through a region of metal which encloses a hole turns out to be affected by the magnetic flux through the hole, even though the magnetic field (which, in classical physics, gives a complete account of all physical effects) vanishes everywhere within the metal itself. In other words, the electrons carrying the current are sensitive to the conditions in a region which they never enter, but only enclose with their paths! This already demonstrates that quantum mechanics forces us to give up some of our classical notions about the 'locality' of physical effects.

A much more severe shock to our conventional picture of the world is administered by Bell's theorem and the related experiments. To describe the upshot in nontechnical terms: given that we make our normal assumptions about local causality in the sense of special relativity theory, and about the statistical properties of ensembles being determined entirely by the initial conditions, then what Bell's theorem and the associated experiments show is that even though two regions of the universe may be spatially separated and physically noninteracting (at least in any sense recognizable to classical physics), we nevertheless cannot ascribe to each of them individual properties; any 'realization' of properties takes place only at the level of the combined system.

Now, it seems to me glaringly obvious – and I am amazed that it does not seem so to all physicists – that the Bell's theorem experiments are a death-knell for reductionism; that Nature is telling us, here, in words of one syllable, that no matter what kinds of map we may build in the future of the physical world, they cannot have built into them the property that the be-

*It is conceivable (though I would deem it unlikely) that recent theoretical work on 'quantum' phase transitions at zero temperature, if experimentally confirmed, might force us to revise this statement.

haviour of the whole is no more than that of the sum of its parts. Of course, for those for whom quantum field theory is the ultimate and only possible map, this is no great surprise, since on inspection it turns out that this particular kind of non-reductionism is inherent from the start in the quantum formalism; and they will no doubt try to dismiss it as in retrospect entirely natural (as Bohr did in his original response to the EPR paradox). But for those who, like me, regard quantum field theory as probably only a transient stage in our understanding of physical reality, Bell's theorem is telling us something very profound and important about the constraints which any future theory must respect. And note, by the way, that it is something which we would not have discovered had we not looked specifically for it.

Let's imagine, though, for the moment, that we take the conventional, conservative view, that is that quantum field theory is indeed the whole truth about the physical world. A defender of this view might, I think, reasonably object that my use of Bell's theorem to argue against reductionism is 'philosophical' (in the sense in which many physicists feel free to dismiss certain kinds of argument as 'merely' philosophical); and in particular that this theorem, while it may no doubt induce us to re-examine exactly what we mean by reductionism in a general sense, gives no reason at all to doubt the particular application of it on which I have fastened above – namely that the behaviour of macroscopic systems is in principle entirely determined by that of their atomic-level constituents. The fact that Bell's theorem has forced us to redefine exactly what we mean by a 'constituent', and in particular to abandon the assumption that it is necessarily associated with a spatially localized region, says nothing against this hypothesis. There is still no reason to doubt – such people would claim – that a solution of the 10^{23}-particle Schrödinger equation for all the electrons and atoms in a macroscopic body, whatever the surprising nonlocal features it might bring out, would nevertheless in principle correctly predict all the properties of that body.

Perhaps not. But I think there is one more feature of our current quantum-mechanical world view which gives us precisely such a reason for doubt – and, I would claim, an overwhelming one: the famous (or infamous!) *quantum measurement paradox. I will now try to review* this in nontechnical language.

Consider an ensemble of systems which can go from some initial to some final state by either of two paths, which we label B and C. (We assume that either or both of these can be shut off if we wish). At the microlevel, we believe – most of us – that despite the fact that 'measurement' of the path followed by any individual system will always show that it followed either B or C, the quantum formalism must nevertheless be interpreted as in some sense saying that if no measurement was made, it simply is not the case that one (unknown) possibility out of B and C was realized; rather, both possibilities are in some sense represented in the correct description. How do we know this? Because, as a matter of experimental fact, the properties of our actual ensemble are not identical to those which we would obtain from a combination of the two ensembles obtained by allowing only B and only C respectively; i.e., we verify, experimentally, the phenomenon of interference between the two paths. So it seems natural to say that the quantum formalism in some sense either ascribes 'reality' to both the possibilities B and C, or ascribes it to neither.

Now suppose that the systems in question are macroscopic and the states B and C macroscopically distinct. Under appropriate circumstances the formalism of quantum mechanics still gives a description – be it by a wave function or a density matrix – in which both possibilities are still represented: in fact, if we extrapolate this formalism up from the microlevel to the macrolevel, there is no point at which any natural discontinuity occurs. In particular, this description automatically ensues if the object in question is to function as a 'measuring apparatus'. What of the interpretation of the formalism? Can we now interpret it as simply saying that any given macrosystem is in state B or state C, with an appropriate probability? At this point there comes in an argument which must have been rehearsed literally thousands of times in the quantum measurement literature, and which is continually rediscovered (and republished!) by those who believe there is really no problem: It is argued that because of the macroscopic nature of the measuring apparatus (or other macro-object involved, e.g. Schrödinger's cat) and/or the fact that it must go from a metastable state to a thermodynamically more stable one and/or its irreducible and irreversible interactions with its environment . . . that there is in fact no possibility of seeing any quantum mechanical interference between the macrostates B and C; all predictions concerning the outcomes of possible experiments will, if properly calculated, be exactly the same for the 'unmeasured' ensemble as for a combination of two ensembles in which *this system was definitely in state B or state C respectively.*

Now I have found by bitter experience that, as regards the majority of adherents of this point of view in the physics community, it is fruitless to argue this point, for the simple reason that they seem unable to conceive the possibility that one might accept – as I do – all the formal statements they make about the lack of observability* of the interference of macrostates, and yet still believe there is a problem. If one thinks that there is an inconsistency in the foundations of quantum measurement theory, it must be because one is ignorant of these elementary and simply demonstrated theorems; and

*In most experimental setups, and in particular in all setups designed as measurement operations in the conventional sense.

so many hours of discussion, not to mention, many pages of Foundations of Physics and other journals, are wasted in telling us, over and over, things which all of us in the other camp already know perfectly well and would not dispute, but which we believe to be totally irrelevant to the real issue.

Why so? The crucial point, I think, is the distinction between the *meaning* of a formalism and the *evidence* that that interpretation of the meaning is correct. At the microlevel, we believe in some sense that the quantum formalism implies that a particular microsystem need not have a particular microstate realized – that both possibilities are in some sense still represented. The *evidence* that this is the correct interpretation of the formalism is the experimentally observed phenomenon of interference. At the macrolevel the formalism of quantum mechanics remains exactly the same; but, for all the reasons cited *ad nauseam* in the literature, there is now no direct experimental evidence against the hypothesis that one of the possibilities B or C has been realized in each particular case. So... can the meaning of the formalism change radically, just because the evidence has disappeared? After all, if in the middle of a murder trial the vital piece of evidence, which would undoubtedly have led to the conviction of the accused, suddenly disappears, he does not thereby become innocent!

In this context I believe that a particular experiment known as the 'macroscopic quantum coherence' (MQC) experiment, which is currently under construction, may be particularly instructive. This is an experiment which is designed inter alia to verify that a genuinely macroscopic object can indeed be in a superposition of macroscopically distinct states as predicted by extrapolation of the formalism of quantum-mechanics. If the outcome of this experiment agrees with these quantum-mechanical predictions, therefore, it will show that a particular kind of macroscopic device (a SQUID ring) does not choose, under the conditions of the experiment, to realize a definite macroscopic state, but that rather, just as at the atomic level, there is in some sense an element of reality associated either with both, or with neither, of the two possible states. Yet this very same device could actually be converted, with apparently trivial changes in the wiring, into a measuring instrument. So – just because we human operators set up a particular piece of metal as a measuring device, does that mean that the system decided to jump out of its quantum-mechanically indefinite state and realize the particular outcome which we observe?

We have here, I think, a clear example of a case in which we have two maps of reality – the quantum-mechanical map which we apply to atomic phenomena, and the 'common-sense', classical map which we use for the macroscopic, everyday world. The problem is that they claim in principle to describe the same level of reality – the world of counters, cats, etc... – and yet no-one has succeeded (at least in my opinion) in showing that they are compatible. Does this matter? That depends, of course, on one's philosophy of science. If one believes that the only goal of a scientific theory is to make correct predictions about the outcome of experiments, then for the moment at least one has no major cause for concern. If on the other hand one requires something more – that vague something which most of us would, I believe, describe as 'understanding' – then one cannot remain satisfied with the current state of affairs, and the obvious corollary is to challenge the whole notion of reductionism.

My own belief – and one which has only strengthened with time – is that the quantum measurement paradox can have no solution within our current reductionist world-view: that it is only a matter of time before advances in technology make this question not merely 'philosophically' but experimentally urgent, and that in the end we will be forced to evolve a picture of the physical world which simply rejects the notion that because quantum-mechanics gives an excellent description of electrons and atoms, and because the macroscopic world is composed (or so we believe!) of electrons and atoms, therefore it follows that quantum mechanics must necessarily and in all respects describe the behaviour of the macroscopic world.

Whether or not this eventually happens will probably not depend on us; how soon it happens may well depend on us, and in particular on us asking the right kinds of question. Right now, the reason we are getting reductionist answers may well be because we are asking only implicitly reductionist questions. The MQC experiment I have mentioned may or may not be the 'right' kind of question; only time will tell. But irrespective of this, I believe that if we think as hard as possible about the implications of quantum mechanics, and take it seriously as applied to the kinds of interestingly complex situations which we can now begin to engineer not only in condensed-matter physics but also in biology and elsewhere, we will more rapidly reach the point where it is seen to break down, and thereby perhaps usher in a radically new picture of physical reality whose nature we cannot at present even guess. I for one intend to use my best efforts to hasten that day.