Antology

BLACK HOLES

and

BABY UNIVERSES

and Other Essays

1993

Stephen Hawking

BANTAM BOOKS
NEW YORK · TORONTO · LONDON · SYDNEY · AUCKLAND

Six

MY POSITION*

THIS ARTICLE IS not about whether I believe in God. Instead, I will discuss my approach to how one can understand the universe: what is the status and meaning of a grand unified theory, a "theory of everything." There is a real problem here. The people who ought to study and argue such questions, the philosophers, have mostly not had enough mathematical background to keep up with modern developments in theoretical physics. There is a subspecies called philosophers of science who ought to be better equipped. But many of them are failed physicists who found it too hard to invent new theories

*Originally given as a talk to a Caius College audience in May 1992.

and so took to writing about the philosophy of physics instead. They are still arguing about the scientific theories of the early years of this century, like relativity and quantum mechanics. They are not in touch with the present frontier of physics.

Maybe I'm being a bit harsh on philosophers, but they have not been very kind to me. My approach has been described as naive and simpleminded. I have been variously called a nominalist, an instrumentalist, a positivist, a realist, and several other ists. The technique seems to be refutation by denigration: If you can attach a label to my approach, you don't have to say what is wrong with it. Surely everyone knows the fatal errors of all those isms.

The people who actually make the advances in theoretical physics don't think in the categories that the philosophers and historians of science subsequently invent for them. I am sure that Einstein, Heisenberg, and Dirac didn't worry about whether they were realists or instrumentalists. They were simply concerned that the existing theories didn't fit together. In theoretical physics, the search for logical self-consistency has always been more important in making advances than experimental results. Otherwise elegant and beautiful theories have been rejected because they don't agree with observation, but I don't know of any major theory that has been advanced just on the basis of experiment. The theory always came first, put forward from the desire to have an elegant and consistent mathematical model. The theory then makes predictions, which can then be tested by observation. If the observations agree with the predictions, that doesn't prove the theory; but the theory survives to make further predictions, which again are tested against observation. If the observations don't agree with the predictions, one abandons the theory.

Or rather, that is what is supposed to happen. In practice, people are very reluctant to give up a theory in which they have

invested a lot of time and effort. They usually start by questioning the accuracy of the observations. If that fails, they try to modify the theory in an ad hoc manner. Eventually the theory becomes a creaking and ugly edifice. Then someone suggests a new theory, in which all the awkward observations are explained in an elegant and natural manner. An example of this was the Michelson-Morley experiment, performed in 1887, which showed that the speed of light was always the same, no matter how the source or the observer was moving. This seemed ridiculous. Surely someone moving toward the light ought to measure it traveling at a higher speed than someone moving in the same direction as the light; yet the experiment showed that both observers would measure exactly the same speed. For the next eighteen years people like Hendrik Lorentz and George Fitzgerald tried to accommodate this observation within accepted ideas of space and time. They introduced ad hoc postulates, such as proposing that objects got shorter when they moved at high speeds. The entire framework of physics became clumsy and ugly. Then in 1905 Einstein suggested a much more attractive viewpoint, in which time was not regarded as completely separate and on its own. Instead it was combined with space in a four-dimensional object called space-time. Einstein was driven to this idea not so much by the experimental results as by the desire to make two parts of the theory fit together in a consistent whole. The two parts were the laws that govern the electric and magnetic fields, and the laws that govern the motion of bodies.

I don't think Einstein, or anyone else in 1905, realized how simple and elegant the new theory of relativity was. It completely revolutionized our notions of space and time. This example illustrates well the difficulty of being a realist in the philosophy of science, for what we regard as reality is conditioned by the theory to which we subscribe. I am certain Lorentz and Fitzgerald regarded themselves as realists, interpreting the experiment on the speed of light in terms of Newtonian ideas of absolute space and absolute time. These notions of space and time seemed to correspond to common sense and reality. Yet nowadays those who are familiar with the theory of relativity, still a disturbingly small minority, have a rather different view. We ought to be telling people about the modern understanding of such basic concepts as space and time.

If what we regard as real depends on our theory, how can we make reality the basis of our philosophy? I would say that I am a realist in the sense that I think there is a universe out there waiting to be investigated and understood. I regard the solipsist position that everything is the creation of our imaginations as a waste of time. No one acts on that basis. But we cannot distinguish what is real about the universe without a theory. I therefore take the view, which has been described as simple-minded or naive, that a theory of physics is just a mathematical model that we use to describe the results of observations. A theory is a good theory if it is an elegant model, if it describes a wide class of observations, and if it predicts the results of new observations. Beyond that, it makes no sense to ask if it corresponds to reality, because we do not know what reality is independent of a theory. This view of scientific theories may make me an instrumentalist or a positivist—as I have said above, I have been called both. The person who called me a positivist went on to add that everyone knew that positivism was out of date—another case of refutation by denigration. It may indeed be out of date in that it was yesterday's intellectual fad, but the positivist position I have outlined seems the only possible one for someone who is seeking new laws, and new ways, to describe the universe. It is no good appealing to reality because we don't have a model independent concept of reality.

In my opinion, the unspoken belief in a model independent reality is the underlying reason for the difficulties philosophers of science have with quantum mechanics and the uncertainty principle. There is a famous thought experiment called Schrödinger's cat. A cat is placed in a sealed box. There is a gun pointing at it, and it will go off if a radioactive nucleus decays. The probability of this happening is fifty percent. (Today no one would dare propose such a thing, even purely as a thought experiment, but in Schrödinger's time they had not heard of animal liberation.)

If one opens the box, one will find the cat either dead or alive. But before the box is opened, the quantum state of the cat will be a mixture of the dead cat state with a state in which the cat is alive. This some philosophers of science find very hard to accept. The cat can't be half shot and half not-shot, they claim, any more than one can be half pregnant. Their difficulty arises because they are implicitly using a classical concept of reality in which an object has a definite single history. The whole point of quantum mechanics is that it has a different view of reality. In this view, an object has not just a single history but all possible histories. In most cases, the probability of having a particular history will cancel out with the probability of having a very slightly different history; but in certain cases, the probabilities of neighboring histories reinforce each other. It is one of these reinforced histories that we observe as the history of the object.

In the case of Schrödinger's cat, there are two histories that are reinforced. In one the cat is shot, while in the other it remains alive. In quantum theory both possibilities can exist together. But some philosophers get themselves tied in knots because they implicitly assume that the cat can have only one history.

The nature of time is another example of an area in which our theories of physics determine our concept of reality. It used to be considered obvious that time flowed on forever, regardless of what was happening; but the theory of relativity combined time with space and said that both could be warped, or distorted, by the matter and energy in the universe. So our perception of the nature of time changed from being independent of the universe to being shaped by it. It then became conceivable that time might simply not be defined before a certain point; as one goes back in time, one might come to an insurmountable barrier, a singularity, beyond which one could not go. If that were the case, it wouldn't make sense to ask who, or what, caused or created the big bang. To talk about causation or creation implicitly assumes there was a time before the big bang singularity. We have known for twenty-five years that Einstein's general theory of relativity predicts that time must have had a beginning in a singularity fifteen billion years ago. But the philosophers have not yet caught up with the idea. They are still worrying about the foundations of quantum mechanics that were laid down sixty-five years ago. They don't realize that the frontier of physics has moved on.

Even worse is the mathematical concept of imaginary time, in which Jim Hartle and I suggested the universe may not have any beginning or end. I was savagely attacked by a philosopher of science for talking about imaginary time. He said: How can a mathematical trick like imaginary time have anything to do with the real universe? I think this philosopher was confusing the technical mathematical terms real and imaginary numbers with the way that real and imaginary are used in everyday language. This just illustrates my point: How can we know what is real, independent of a theory or model with which to interpret it?

I have used examples from relativity and quantum mechanics to show the problems one faces when one tries to make sense of the universe. It doesn't really matter if you don't un-

derstand relativity and quantum mechanics, or even if these theories are incorrect. What I hope I have demonstrated is that some sort of positivist approach, in which one regards a theory as a model, is the only way to understand the universe, at least for a theoretical physicist. I am hopeful that we will find a consistent model that describes everything in the universe. If we do that, it will be a real triumph for the human race.

Twelve

IS EVERYTHING DETERMINED?*

IN THE PLAY *Julius Caesar*, Cassius tells Brutus, "Men at some times are masters of their fate." But are we really masters of our fate? Or is everything we do determined and preordained? The argument for preordination used to be that God was omnipotent and outside time, so God would know what was going to happen. But how then could we have any free will? And if we don't have free will, how can we be responsible for our actions? It can hardly be one's fault if one has been preordained to rob a bank. So why should one be punished for it?

*A lecture given at the Sigma Club seminar at the University of Cambridge, April 1990.

In recent times, the argument for determinism has been based on science. It seems that there are well-defined laws that govern how the universe and everything in it develops in time. Although we have not yet found the exact form of all these laws, we already know enough to determine what happens in all but the most extreme situations. Whether we will find the remaining laws in the fairly near future is a matter of opinion. I'm an optimist: I think there's a fifty-fifty chance that we will find them in the next twenty years. But even if we don't, it won't really make any difference to the argument. The important point is that there should exist a set of laws that completely determines the evolution of the universe from its initial state. These laws may have been ordained by God. But it seems that He (or She) does not intervene in the universe to break the laws.

The initial configuration of the universe may have been chosen by God, or it may itself have been determined by the laws of science. In either case, it would seem that everything in the universe would then be determined by evolution according to the laws of science, so it is difficult to see how we can be masters of our fate.

The idea that there is some grand unified theory that determines everything in the universe raises many difficulties. First of all, the grand unified theory is presumably compact and elegant in mathematical terms. There ought to be something special and simple about the theory of everything. Yet how can a certain number of equations account for the complexity and trivial detail that we see around us? Can one really believe that the grand unified theory has determined that Sinead O'Connor will be the top of the hit parade this week, or that Madonna will be on the cover of *Cosmopolitan*?

A second problem with the idea that everything is determined by a grand unified theory is that anything we say is also

determined by the theory. But why should it be determined to be correct? Isn't it more likely to be wrong, because there are many possible incorrect statements for every true one? Each week, my mail contains a number of theories that people have sent me. They are all different, and most are mutually inconsistent. Yet presumably the grand unified theory has determined that the authors think they were correct. So why should anything I say have any greater validity? Aren't I equally determined by the grand unified theory?

A third problem with the idea that everything is determined is that we feel that we have free will—that we have the freedom to choose whether to do something. But if everything is determined by the laws of science, then free will must be an illusion, and if we don't have free will, what is the basis for our responsibility for our actions? We don't punish people for crimes if they are insane, because we have decided that they can't help it. But if we are all determined by a grand unified theory, none of us can help what we do, so why should anyone be held responsible for what they do?

These problems of determinism have been discussed over the centuries. The discussion was somewhat academic, however, as we were far from a complete knowledge of the laws of science, and we didn't know how the initial state of the universe was determined. The problems are more urgent now because there is the possibility that we may find a complete unified theory in as little as twenty years. And we realize that the initial state may itself have been determined by the laws of science. What follows is my personal attempt to come to terms with these problems. I don't claim any great originality or depth, but it is the best I can do at the moment.

To start with the first problem: How can a relatively simple and compact theory give rise to a universe that is as complex as the one we observe, with all its trivial and unimportant details? The key to this is the uncertainty principle of quantum mechanics, which states that one cannot measure both the position and speed of a particle to great accuracy; the more accurately you measure the position, the less accurately you can measure the speed, and vice versa. This uncertainty is not so important at the present time, when things are far apart, so that a small uncertainty in position does not make much difference. But in the very early universe, everything was very close together, so there was quite a lot of uncertainty, and there were a number of possible states for the universe. These different possible early states would have evolved into a whole family of different histories for the universe. Most of these histories would be similar in their large-scale features. They would correspond to a universe that was uniform and smooth, and that was expanding. However, they would differ on details like the distribution of stars and, even more, on what was on the covers of their magazines. (That is, if those histories contained magazines.) Thus the complexity of the universe around us and its details arose from the uncertainty principle in the early stages. This gives a whole family of possible histories for the universe. There would be a history in which the Nazis won the Second World War, though the probability is low. But we just happen to live in a history in which the Allies won the war and Madonna was on the cover of *Cosmopolitan*.

I now turn to the second problem: If what we do is determined by some grand unified theory, why should the theory determine that we draw the right conclusions about the universe rather than the wrong ones? Why should anything we say have any validity? My answer to this is based on Darwin's idea of natural selection. I take it that some

very primitive form of life arose spontaneously on earth from chance combinations of atoms. This early form of life was probably a large molecule. But it was probably not DNA, since the chances of forming a whole DNA molecule by random combinations are small.

The early form of life would have reproduced itself. The quantum uncertainty principle and the random thermal motions of the atoms would mean that there were a certain number of errors in the reproduction. Most of these errors would have been fatal to the survival of the organism or its ability to reproduce. Such errors would not be passed on to future generations but would die out. A very few errors would be beneficial, by pure chance. The organisms with these errors would be more likely to survive and reproduce. Thus they would tend to replace the original, unimproved organisms.

The development of the double helix structure of DNA may have been one such improvement in the early stages. This was probably such an advance that it completely replaced any earlier form of life, whatever that may have been. As evolution progressed, it would have led to the development of the central nervous system. Creatures that correctly recognized the implications of data gathered by their sense organs and took appropriate action would be more likely to survive and reproduce. The human race has carried this to another stage. We are very similar to higher apes, both in our bodies and in our DNA; but a slight variation in our DNA has enabled us to develop language. This has meant that we can hand down information and accumulated experience from generation to generation, in spoken and eventually in written form. Previously, the results of experience could be handed down only by the slow process of it being encoded into DNA through random errors in reproduction. The effect has been a dramatic speed-up of evolution.

It took more than three billion years to evolve up to the human race. But in the course of the last ten thousand years, we have developed written language. This has enabled us to progress from cave dwellers to the point where we can ask about the ultimate theory of the universe.

There has been no significant biological evolution, or change in human DNA, in the last ten thousand years. Thus, our intelligence, our ability to draw the correct conclusions from the information provided by our sense organs, must date back to our cave dweller days or earlier. It would have been selected for on the basis of our ability to kill certain animals for food and to avoid being killed by other animals. It is remarkable that mental qualities that were selected for these purposes should have stood us in such good stead in the very different circumstances of the present day. There is probably not much survival advantage to be gained from discovering a grand unified theory or answering questions about determinism. Nevertheless, the intelligence that we have developed for other reasons may well ensure that we find the right answers to these questions.

I now turn to the third problem, the questions of free will and responsibility for our actions. We feel subjectively that we have the ability to choose who we are and what we do. But this may just be an illusion. Some people think they are Jesus Christ or Napoleon, but they can't all be right. What we need is an objective test that we can apply from the outside to distinguish whether an organism has free will. For example, suppose we were visited by a "little green person" from another star. How could we decide whether it had free will or was just a robot, programmed to respond as if it were like us?

The ultimate objective test of free will would seem to

be: Can one predict the behavior of the organism? If one can, then it clearly doesn't have free will but is predetermined. On the other hand, if one cannot predict the behavior, one could take that as an operational definition that the organism has free will.

One might object to this definition of free will on the grounds that once we find a complete unified theory we will be able to predict what people will do. The human brain, however, is also subject to the uncertainty principle. Thus, there is an element of the randomness associated with quantum mechanics in human behavior. But the energies involved in the brain are low, so quantum mechanical uncertainty is only a small effect. The real reason why we cannot predict human behavior is that it is just too difficult. We already know the basic physical laws that govern the activity of the brain, and they are comparatively simple. But it is just too hard to solve the equations when there are more than a few particles involved. Even in the simpler Newtonian theory of gravity, one can solve the equations exactly only in the case of two particles. For three or more particles one has to resort to approximations, and the difficulty increases rapidly with the number of particles. The human brain contains about 10^{26} or a hundred million billion billion particles. This is far too many for us ever to be able to solve the equations and predict how the brain would behave, given its initial state and the nerve data coming into it. In fact, of course, we cannot even measure what the initial state was, because to do so we would have to take the brain apart. Even if we were prepared to do that, there would just be too many particles to record. Also, the brain is probably very sensitive to the initial state—a small change in the initial state can make a very large difference to subsequent behavior. So although we know the fundamental equations that govern the brain, we are quite unable to use them to predict human behavior.

This situation arises in science whenever we deal with the macroscopic system, because the number of particles is always too large for there to be any chance of solving the fundamental equations. What we do instead is use effective theories. These are approximations in which the very large number of particles are replaced by a few quantities. An example is fluid mechanics. A liquid such as water is made up of billions of billions of molecules that themselves are made up of electrons, protons, and neutrons. Yet it is a good approximation to treat the liquid as a continuous medium, characterized just by velocity, density, and temperature. The predictions of the effective theory of fluid mechanics are not exact—one only has to listen to the weather forecast to realize that—but they are good enough for the design of ships or oil pipelines.

I want to suggest that the concepts of free will and moral responsibility for our actions are really an effective theory in the sense of fluid mechanics. It may be that everything we do is determined by some grand unified theory. If that theory has determined that we shall die by hanging, then we shall not drown. But you would have to be awfully sure that you were destined for the gallows to put to sea in a small boat during a storm. I have noticed that even people who claim that everything is predestined and that we can do nothing to change it look before they cross the road. Maybe it's just that those who don't look don't survive to tell the tale.

One cannot base one's conduct on the idea that everything is determined, because one does not know what has been determined. Instead, one has to adopt the effective theory that one has free will and that one is responsible for one's actions. This theory is not very good at predicting human behavior, but we adopt it because there is no chance of solving the equations arising from the fundamental laws. There is also a Darwinian reason that we believe in free will: A society in which the individual feels responsible for his or her actions is more likely to work together and survive to spread its values. Of course, ants work well together. But such a society is static. It cannot respond to unfamiliar challenges or develop new opportunities. A collection of free individuals who share certain mutual aims, however, can collaborate on their common objectives and yet have the flexibility to make innovations. Thus, such a society is more likely to prosper and to spread its system of values.

The concept of free will belongs to a different arena from that of fundamental laws of science. If one tries to deduce human behavior from the laws of science, one gets caught in the logical paradox of self-referencing systems. If what one does could be predicted from the fundamental laws, then the fact of making that prediction could change what happens. It is like the problems one would get into if time travel were possible, which I don't think it ever will be. If you could see what is going to happen in the future, you could change it. If you knew which horse was going to win the Grand National, you could make a fortune by betting on it. But that action would change the odds. One only has to see *Back to the Future* to realize what problems could arise.

This paradox about being able to predict one's actions is closely related to the problem I mentioned earlier: Will the ultimate theory determine that we come to the right conclusions about the ultimate theory? In that case, I argued that Darwin's idea of natural selection would lead us to the correct answer. Maybe the correct answer is not the right way to describe it, but natural selection should at least lead us to a set of physical laws that work fairly well. However, we cannot apply those physical laws to deduce human behavior for two reasons. First, we cannot solve the equations. Second, even if we could, the fact of making a prediction would disturb the system. Instead, natural selection seems to lead to us adopting the effective theory of free will. If one accepts that a person's actions are freely chosen, one cannot then argue that in some cases they are determined by outside forces. The concept of "almost free will" doesn't make sense. But people tend to confuse the fact that one may be able to guess what an individual is likely to choose with the notion that the choice is not free. I would guess that most of you will have a meal this evening, but you are quite free to choose to go to bed hungry. One example of such confusion is the doctrine of diminished responsibility: the idea that persons should not be punished for their actions because they were under stress. It may be that someone is more likely to commit an antisocial act when under stress. But that does not mean that we should make it even more likely that he or she commit the act by reducing the punishment.

One has to keep the investigation of the fundamental laws of science and the study of human behavior in separate compartments. One cannot use the fundamental laws to deduce human behavior, for the reasons I have explained. But one might hope that we could employ both the intelligence and the powers of logical thought that we have developed through natural selection. Unfortunately, natural selection has also developed other characteristics, such as aggression. Aggression would have given a survival advantage in cave dweller days and earlier and so would have been favored by natural selection. The tremendous increase in our powers of destruction brought about by modern science and technology, however, has made aggression a very dangerous quality, one that threat-

ens the survival of the whole human race. The trouble is, our aggressive instincts seem to be encoded in our DNA. DNA changes by biological evolution only on a time scale of millions of years, but our powers of destruction are increasing on a time scale for the evolution of information, which is now only twenty or thirty years. Unless we can use our intelligence to control our aggression, there is not much chance for the human race. Still, while there's life, there's hope. If we can survive the next hundred years or so, we will have spread to other planets and possibly to other stars. This will make it much less likely that the entire human race will be wiped out by a calamity such as a nuclear war.

To recapitulate: I have discussed some of the problems that arise if one believes that everything in the universe is determined. It doesn't make much difference whether this determinism is due to an omnipotent God or to the laws of science. Indeed, one could always say that the laws of science are the expression of the will of God.

I considered three questions: First, how can the complexity of the universe and all its trivial details be determined by a simple set of equations? Alternatively, can one really believe that God chose all the trivial details, like who should be on the cover of *Cosmopolitan*? The answer seems to be that the uncertainty principle of quantum mechanics means that there is not just a single history for the universe but a whole family of possible histories. These histories may be similar on very large scales, but they will differ greatly on normal, everyday scales. We happen to live on one particular history that has certain properties and details. But there are very similar intelligent beings who live on histories that differ in who won the war and who is Top of the Pops. Thus, the trivial details of our universe arise because the fundamental laws incorporate quantum mechanics with its element of uncertainty or randomness.

The second question was: If everything is determined by some fundamental theory, then what we say about the theory is also determined by the theory—and why should it be determined to be correct, rather than just plain wrong or irrelevant? My answer to this was to appeal to Darwin's theory of natural selection: Only those individuals who drew the appropriate conclusions about the world around them would be likely to survive and reproduce.

The third question was: If everything is determined, what becomes of free will and our responsibility for our actions? But the only objective test of whether an organism has free will is whether its behavior can be predicted. In the case of human beings, we are quite unable to use the fundamental laws to predict what people will do, for two reasons. First, we cannot solve the equations for the very large number of particles involved. Second, even if we could solve the equations, the fact of making a prediction would disturb the system and could lead to a different outcome. So as we cannot predict human behavior, we may as well adopt the effective theory that humans are free agents who can choose what to do. It seems that there are definite survival advantages to believing in free will and responsibility for one's actions. That means this belief should be reinforced by natural selection. Whether the language-transmitted sense of responsibility is sufficient to control the DNA-transmitted instinct of aggression remains to be seen. If it does not, the human race will have been one of natural selection's dead ends. Maybe some other race of intelligent beings elsewhere in the galaxy will achieve a better balance between responsibility and aggression. But if so, we might have expected to be contacted by them, or at least to detect their radio signals. Maybe they are aware of our existence but don't want

to reveal themselves to us. That might be wise, given our record.

In summary, the title of this essay was a question: Is everything determined? The answer is yes, it is. But it might as well not be, because we can never know what is determined.