

## CHAPTER 11

## THE UNIFICATION OF PHYSICS

As was explained in the first chapter, it would be very difficult to construct a complete unified theory of everything in the universe all at one go. So instead we have made progress by finding partial theories that describe a limited range of happenings and by neglecting other effects or approximating them by certain numbers. (Chemistry, for example, allows us to calculate the interactions of atoms, without knowing the internal structure of an atom's nucleus.) Ultimately, however, one would hope to find a complete, consistent, unified theory that would include all these partial theories as approximations, and that did not need to be adjusted to fit the facts by picking the values of certain arbitrary numbers in the theory. The quest for such a theory is known as "the unification of physics." Einstein spent most of his later years unsuccessfully searching for a unified theory, but the time was not ripe: there were partial theories for gravity and the electromagnetic force, but very little was known about the nuclear forces. Moreover, Einstein refused to believe in the reality of quantum mechanics, despite the important role he had played in its development. Yet it seems that the uncertainty principle is a fundamental feature of the universe we live in. A successful unified theory must, therefore, necessarily incorporate this principle.

As I shall describe, the prospects for finding such a theory seem to be much better now because we know so much more about the universe. But we must beware of overconfidence – we have had false dawns before! At the beginning of this century, for example, it was thought that everything could be explained in terms of the properties of continuous matter, such as elasticity and heat conduction. The discovery of atomic structure and the uncertainty principle put an emphatic end to that. Then again, in 1928, physicist and Nobel Prize winner Max Born told a group of visitors to Gottingen University, "Physics, as we know it, will be over in six months." His confidence was based on the recent discovery by Dirac of the equation that governed the electron. It was thought that a similar equation would govern the proton, which was the only other particle known at the time, and that would be the end of theoretical physics. However, the discovery of the neutron and of nuclear forces knocked that one on the head too. Having said this, I still believe there are grounds for cautious optimism that we may now be near the end of the search for the ultimate laws of nature.

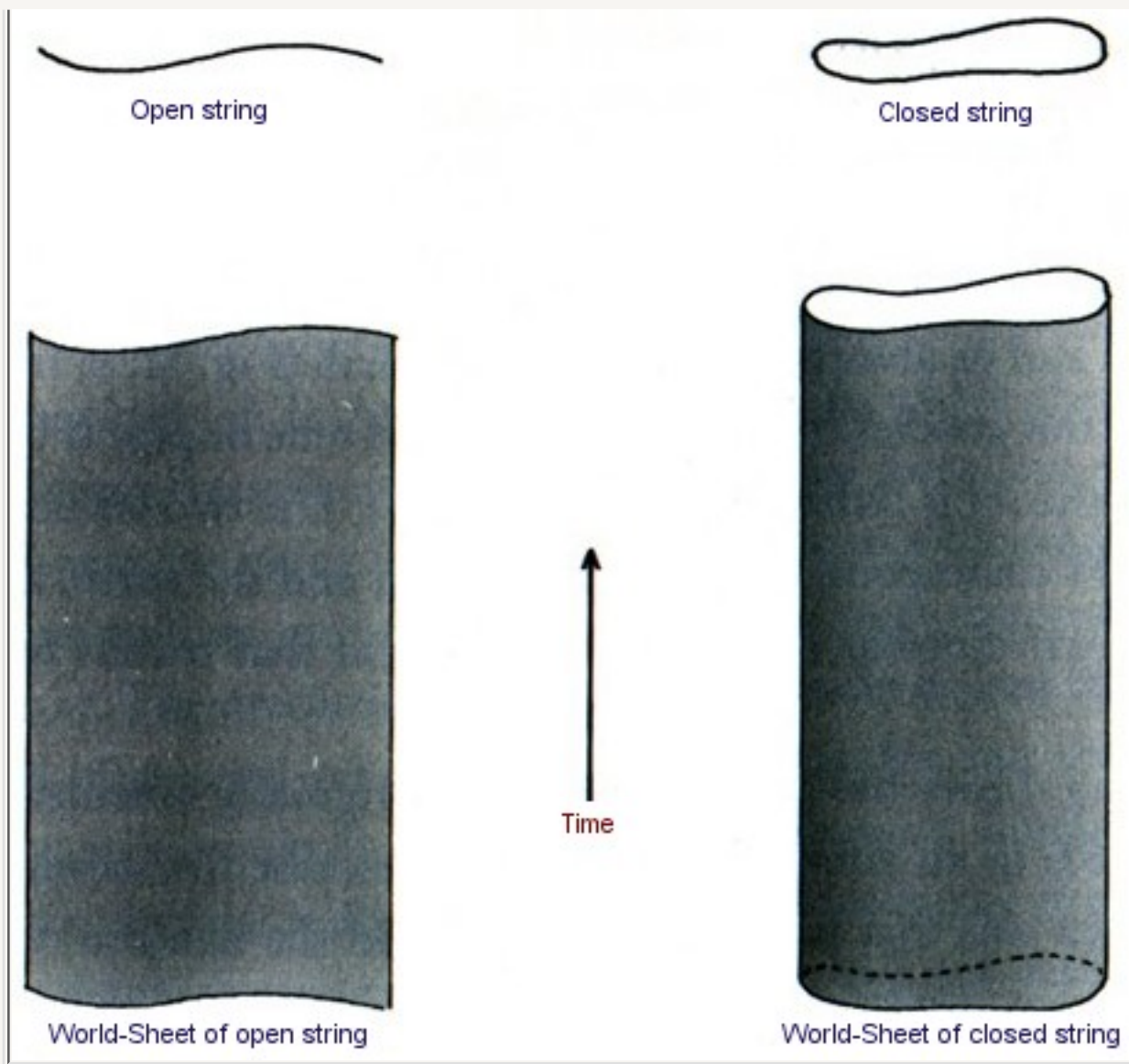
In previous chapters I have described general relativity, the partial theory of gravity, and the partial theories that govern the weak, the strong, and the electromagnetic forces. The last three may be combined in so-called grand unified theories, or GUTs, which are not very satisfactory because they do not include gravity and because they contain a number of quantities, like the relative masses of different particles, that cannot be predicted from the theory but have to be chosen to fit observations. The main difficulty in finding a theory that unifies gravity with the other forces is that general relativity is a "classical" theory; that is, it does not incorporate the uncertainty principle of quantum mechanics. On the other hand, the other partial theories depend on quantum mechanics in an essential way. A necessary first step, therefore, is to combine general relativity with the uncertainty principle. As we have seen, this can produce some remarkable consequences, such as black holes not being black, and the universe not having any singularities but being completely self-contained and without a boundary. The trouble is, as explained in Chapter 7, that the uncertainty principle means that even "empty" space is filled with pairs of virtual particles and antiparticles. These pairs would have an infinite amount of energy and, therefore, by Einstein's famous equation  $E = mc^2$ , they would have an infinite amount of mass. Their gravitational attraction would thus curve up the universe to infinitely small size.

Rather similar, seemingly absurd infinities occur in the other partial theories, but in all these cases the infinities can be canceled out by a process called renormalization. This involves canceling the infinities by introducing other infinities. Although this technique is rather dubious mathematically, it does seem to work in practice, and has been used with these theories to make predictions that agree with observations to an extraordinary degree of accuracy. Renormalization, however, does have a serious drawback from the point of view of trying to find a complete theory, because it means that the actual values of the masses and the strengths of the forces cannot be predicted from the theory, but have to be chosen to fit the observations.

In attempting to incorporate the uncertainty principle into general relativity, one has only two quantities that can be adjusted: the strength of gravity and the value of the cosmological constant. But adjusting these is not

sufficient to remove all the infinities. One therefore has a theory that seems to predict that certain quantities, such as the curvature of space-time, are really infinite, yet these quantities can be observed and measured to be perfectly finite! This problem in combining general relativity and the uncertainty principle had been suspected for some time, but was finally confirmed by detailed calculations in 1972. Four years later, a possible solution, called “supergravity,” was suggested. The idea was to combine the spin-2 particle called the graviton, which carries the gravitational force, with certain other particles of spin  $3/2$ ,  $1$ ,  $1/2$ , and  $0$ . In a sense, all these particles could then be regarded as different aspects of the same “superparticle,” thus unifying the matter particles with spin  $1/2$  and  $3/2$  with the force-carrying particles of spin  $0$ ,  $1$ , and  $2$ . The virtual particle/antiparticle pairs of spin  $1/2$  and  $3/2$  would have negative energy, and so would tend to cancel out the positive energy of the spin  $2$ ,  $1$ , and  $0$  virtual pairs. This would cause many of the possible infinities to cancel out, but it was suspected that some infinities might still remain. However, the calculations required to find out whether or not there were any infinities left uncanceled were so long and difficult that no one was prepared to undertake them. Even with a computer it was reckoned it would take at least four years, and the chances were very high that one would make at least one mistake, probably more. So one would know one had the right answer only if someone else repeated the calculation and got the same answer, and that did not seem very likely!

Despite these problems, and the fact that the particles in the super-gravity theories did not seem to match the observed particles, most scientists believed that supergravity was probably the right answer to the problem of the unification of physics. It seemed the best way of unifying gravity with the other forces. However, in 1984 there was a remarkable change of opinion in favor of what are called string theories. In these theories the basic objects are not particles, which occupy a single point of space, but things that have a length but no other dimension, like an infinitely thin piece of string. These strings may have ends (the so-called open strings) or they may be joined up with themselves in closed loops (closed strings) Figure 11:1 and Figure 11:2.



Figures 11:1 &amp; 11:2

A particle occupies one point of space at each instant of time. Thus its history can be represented by a line in space-time (the “world-line”). A string, on the other hand, occupies a line in space at each moment of time. So its history in space-time is a two-dimensional surface called the world-sheet. (Any point on such a world-sheet can be described by two numbers, one specifying the time and the other the position of the point on the string.) The world-sheet of an open string is a strip: its edges represent the paths through space-time of the ends of the string **Figure 11:1**. The world-sheet of a closed string is a cylinder or tube **Figure 11:2**: a slice through the tube is a circle, which represents the position of the string at one particular time.

Two pieces of string can join together to form a single string; in the case of open strings they simply join at the ends **Figure 11:3**, while in the case of closed strings it is like the two legs joining on a pair of trousers **Figure 11:4**.

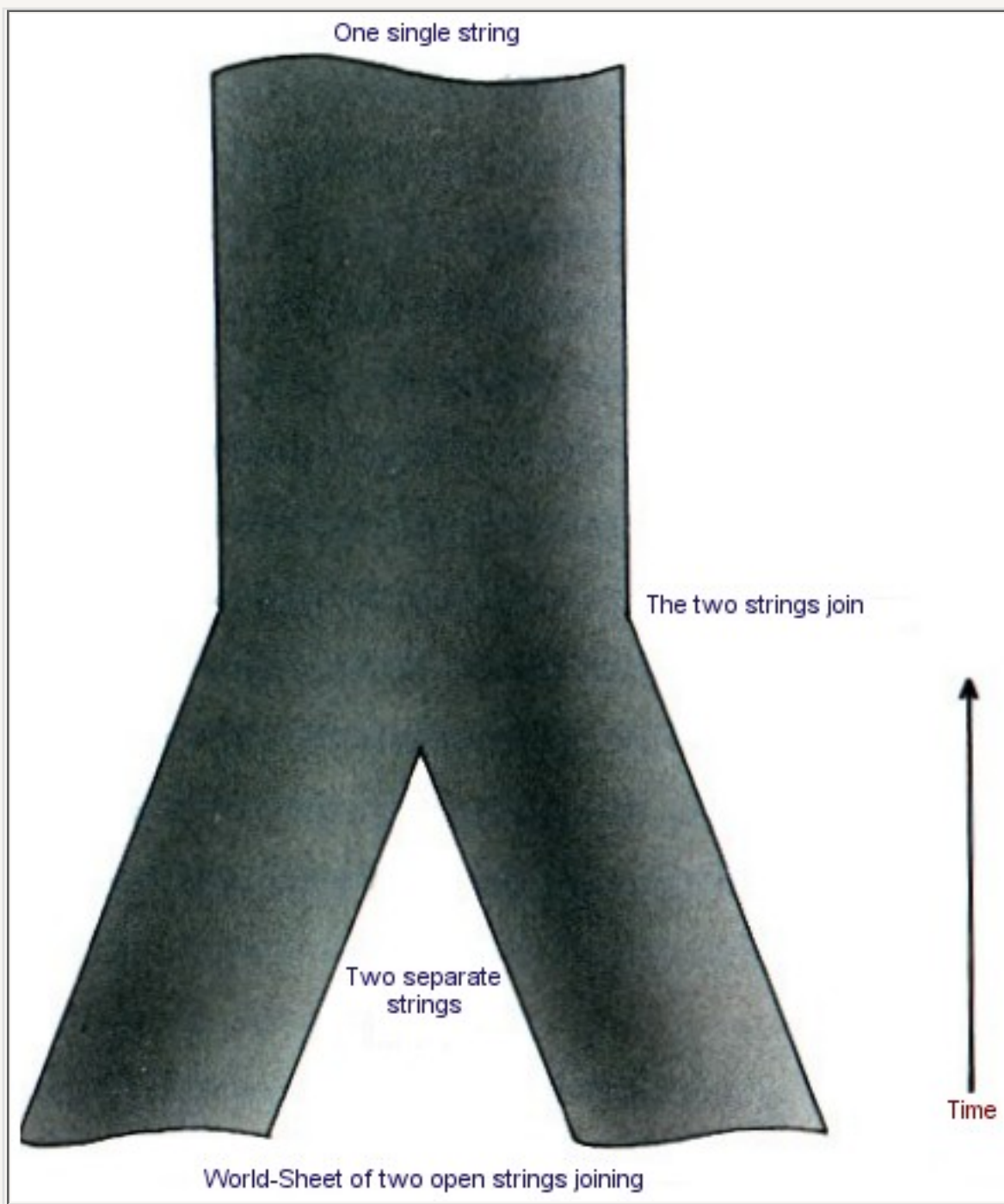


Figure 11:3



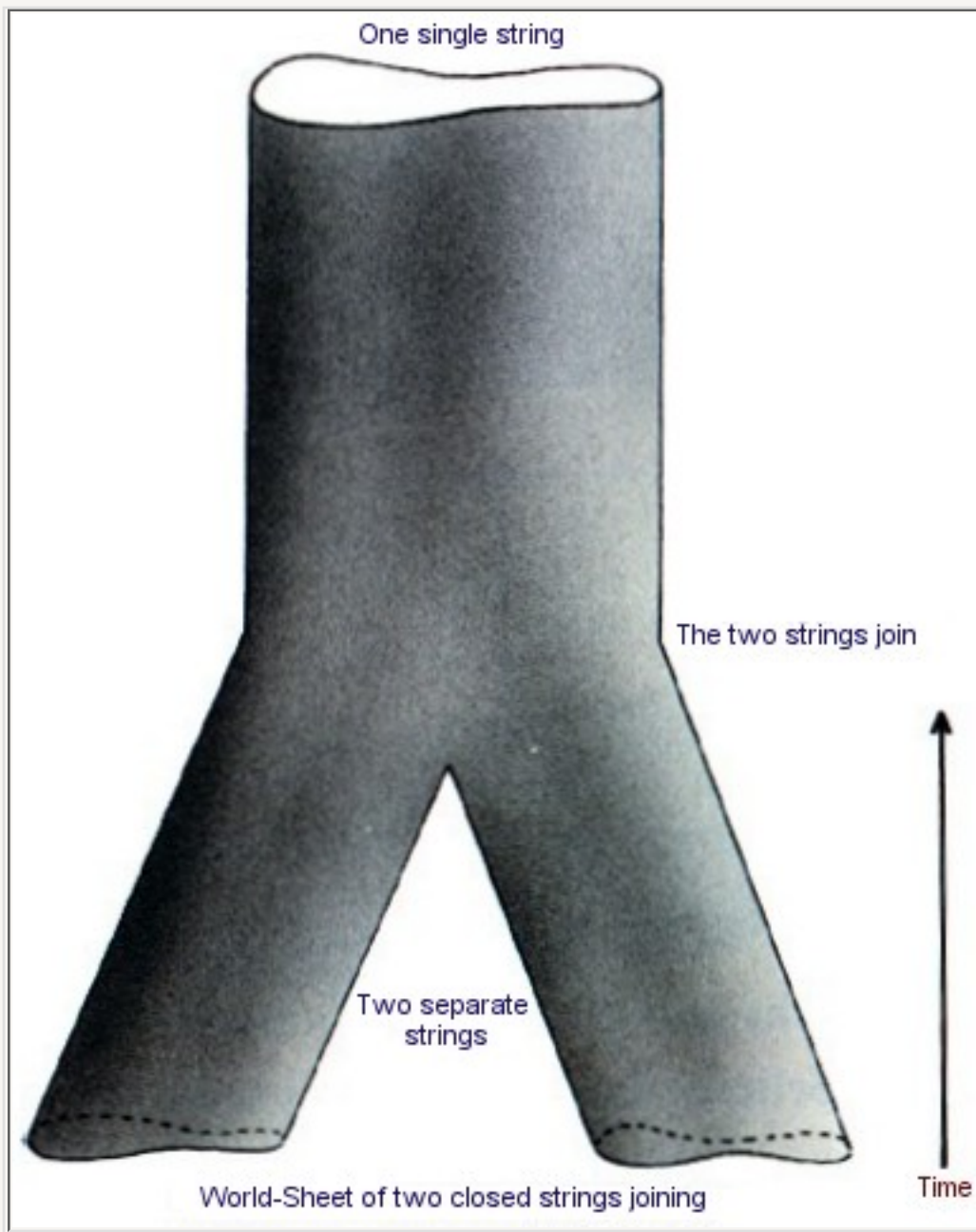
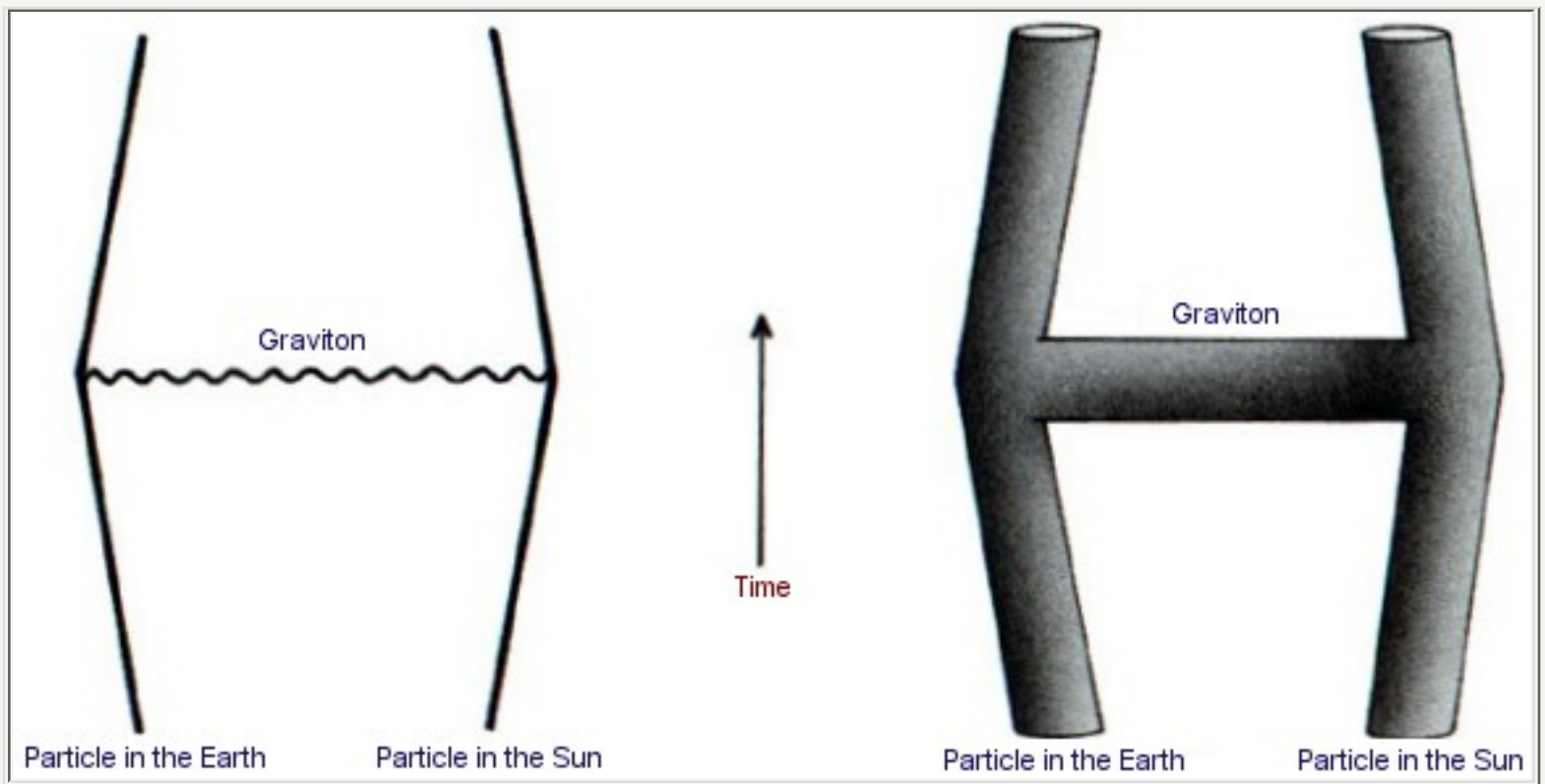


Figure 11:4

Similarly, a single piece of string can divide into two strings. In string theories, what were previously thought of as particles are now pictured as waves traveling down the string, like waves on a vibrating kite string. The emission or absorption of one particle by another corresponds to the dividing or joining together of strings. For example, the gravitational force of the sun on the earth was pictured in particle theories as being caused by the emission of a graviton by a particle in the sun and its absorption by a particle in the earth Figure 11:5.



Figures 11:5 & 11:6

In string theory, this process corresponds to an H-shaped tube or pipe **Figure 11:6** (string theory is rather like plumbing, in a way). The two vertical sides of the H correspond to the particles in the sun and the earth, and the horizontal crossbar corresponds to the graviton that travels between them.

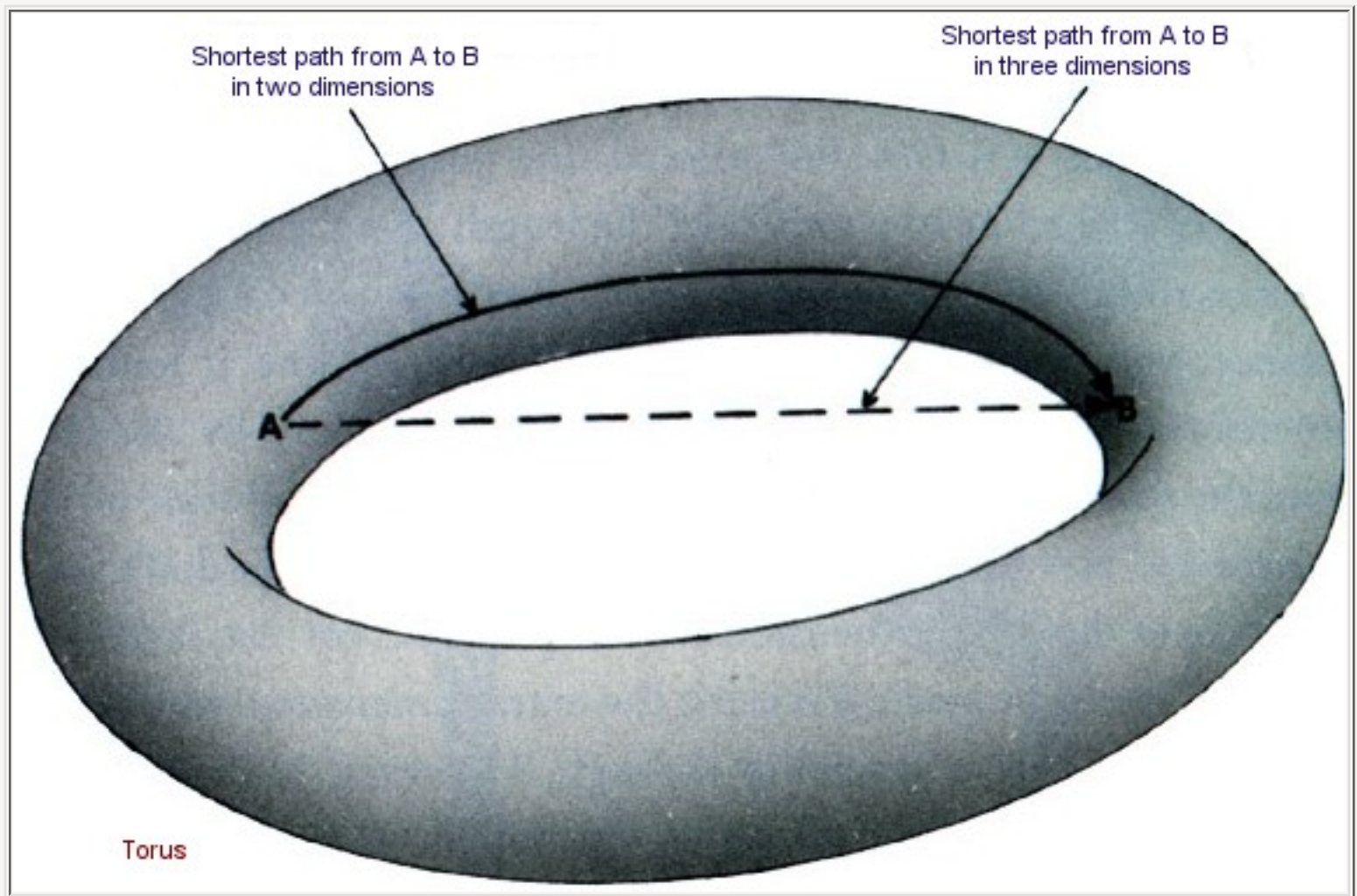
String theory has a curious history. It was originally invented in the late 1960s in an attempt to find a theory to describe the strong force. The idea was that particles like the proton and the neutron could be regarded as waves on a string. The strong forces between the particles would correspond to pieces of string that went between other bits of string, as in a spider's web. For this theory to give the observed value of the strong force between particles, the strings had to be like rubber bands with a pull of about ten tons.

In 1974 Joel Scherk from Paris and John Schwarz from the California Institute of Technology published a paper in which they showed that string theory could describe the gravitational force, but only if the tension in the string were very much higher, about a thousand million million million million million tons (1 with thirty-nine zeros after it). The predictions of the string theory would be just the same as those of general relativity on normal length scales, but they would differ at very small distances, less than a thousand million million million million millionth of a centimeter (a centimeter divided by 1 with thirty-three zeros after it). Their work did not receive much attention, however, because at just about that time most people abandoned the original string theory of the strong force in favor of the theory based on quarks and gluons, which seemed to fit much better with observations. Scherk died in tragic circumstances (he suffered from diabetes and went into a coma when no one was around to give him an injection of insulin). So Schwarz was left alone as almost the only supporter of string theory, but now with the much higher proposed value of the string tension.

In 1984 interest in strings suddenly revived, apparently for two reasons. One was that people were not really making much progress toward showing that supergravity was finite or that it could explain the kinds of particles that we observe. The other was the publication of a paper by John Schwarz and Mike Green of Queen Mary College, London, that showed that string theory might be able to explain the existence of particles that have a built-in left-handedness, like some of the particles that we observe. Whatever the reasons, a large number of people soon began to work on string theory and a new version was developed, the so-called heterotic string, which seemed as if it might be able to explain the types of particles that we observe.

String theories also lead to infinities, but it is thought they will all cancel out in versions like the heterotic string

(though this is not yet known for certain). String theories, however, have a bigger problem: they seem to be consistent only if space-time has either ten or twenty-six dimensions, instead of the usual four! Of course, extra space-time dimensions are a commonplace of science fiction indeed, they provide an ideal way of overcoming the normal restriction of general relativity that one cannot travel faster than light or back in time (see Chapter 10). The idea is to take a shortcut through the extra dimensions. One can picture this in the following way. Imagine that the space we live in has only two dimensions and is curved like the surface of an anchor ring or torus **Figure 11:7**.



**Figure 11:7**

If you were on one side of the inside edge of the ring and you wanted to get to a point on the other side, you would have to go round the inner edge of the ring. However, if you were able to travel in the third dimension, you could cut straight across.

Why don't we notice all these extra dimensions, if they are really there? Why do we see only three space dimensions and one time dimension? The suggestion is that the other dimensions are curved up into a space of very small size, something like a million million million million millionth of an inch. This is so small that we just don't notice it: we see only one time dimension and three space dimensions, in which space-time is fairly flat. It is like the surface of a straw. If you look at it closely, you see it is two-dimensional (the position of a point on the straw is described by two numbers, the length along the straw and the distance round the circular direction). But if you look at it from a distance, you don't see the thickness of the straw and it looks one-dimensional (the position of a point is specified only by the length along the straw). So it is with space-time: on a very small scale it is ten-dimensional and highly curved, but on bigger scales you don't see the curvature or the extra dimensions. If this picture is correct, it spells bad news for would-be space travelers: the extra dimensions would be far too small to allow a spaceship through. However, it raises another major problem. Why should some, but not all, of the dimensions be curled up into a small ball? Presumably, in the very early universe all the dimensions would have been very curved. Why did one time dimension and three space dimensions flatten



out, while the other dimensions remain tightly curled up?

One possible answer is the anthropic principle. Two space dimensions do not seem to be enough to allow for the development of complicated beings like us. For example, two-dimensional animals living on a one-dimensional earth would have to climb over each other in order to get past each other. If a two-dimensional creature ate something it could not digest completely, it would have to bring up the remains the same way it swallowed them, because if there were a passage right through its body, it would divide the creature into two separate halves: our two-dimensional being would fall apart. Figure 11:8. Similarly, it is difficult to see how there could be any circulation of the blood in a two-dimensional creature.

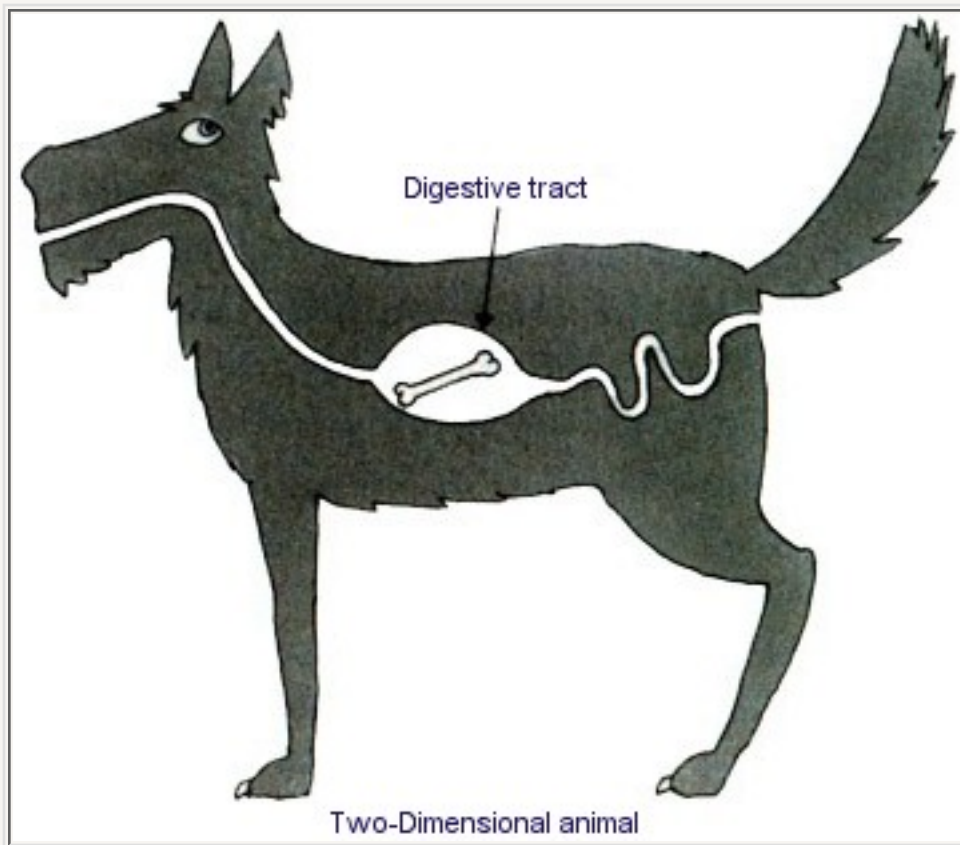


Figure 11:8

There would also be problems with more than three space dimensions. The gravitational force between two bodies would decrease more rapidly with distance than it does in three dimensions. (In three dimensions, the gravitational force drops to  $1/4$  if one doubles the distance. In four dimensions it would drop to  $1/5$ , in five dimensions to  $1/6$ , and so on.) The significance of this is that the orbits of planets, like the earth, around the sun would be unstable: the least disturbance from a circular orbit (such as would be caused by the gravitational attraction of other planets) would result in the earth spiraling away from or into the sun. We would either freeze or be burned up. In fact, the same behavior of gravity with distance in more than three space dimensions means that the sun would not be able to exist in a stable state with pressure balancing gravity. It would either fall apart or it would collapse to form a black hole. In either case, it would not be of much use as a source of heat and light for life on earth. On a smaller scale, the electrical forces that cause the electrons to orbit round the nucleus in an atom would behave in the same way as gravitational forces. Thus the electrons would either escape from the atom altogether or would spiral into the nucleus. In either case, one could not have atoms as we know them.

It seems clear then that life, at least as we know it, can exist only in regions of space-time in which one time dimension and three space dimensions are not curled up small. This would mean that one could appeal to the weak anthropic principle, provided one could show that string theory does at least allow there to be such regions of the universe – and it seems that indeed string theory does. There may well be other regions of the universe, or other universes (whatever *that* may mean), in which all the dimensions are curled up small or in which more than four dimensions are nearly flat, but there would be no intelligent beings in such regions to



observe the different number of effective dimensions.

Another problem is that there are at least four different string theories (open strings and three different closed string theories) and millions of ways in which the extra dimensions predicted by string theory could be curled up. Why should just one string theory and one kind of curling up be picked out? For a time there seemed no answer, and progress got bogged down. Then, from about 1994, people started discovering what are called **dualities**: different string theories and different ways of curling up the extra dimensions could lead to the same results in four dimensions. Moreover, as well as particles, which occupy a single point of space, and strings, which are lines, there were found to be other objects called **p-branes**, which occupied two-dimensional or higher-dimensional volumes in space. (A particle can be regarded as a 0-brane and a string as a 1-brane but there were also p-branes for  $p=2$  to  $p=9$ .) What this seems to indicate is that there is a sort of democracy among supergravity, string, and p-brane theories: they seem to fit together but none can be said to be more fundamental than the others. They appear to be different approximations to some fundamental theory that are valid in different situations.

People have searched for this underlying theory, but without any success so far. However, I believe there may not be any single formulation of the fundamental theory any more than, as Godel showed, one could formulate arithmetic in terms of a single set of axioms. Instead it may be like maps – you can't use a single map to describe the surface of the earth or an anchor ring: you need at least two maps in the case of the earth and four for the anchor ring to cover every point. Each map is valid only in a limited region, but different maps will have a region of overlap. The collection of maps provides a complete description of the surface. Similarly, in physics it may be necessary to use different formulations in different situations, but two different formulations would agree in situations where they can both be applied. The whole collection of different formulations could be regarded as a complete unified theory, though one that could not be expressed in terms of a single set of postulates.

But can there really be such a unified theory? Or are we perhaps just chasing a mirage? There seem to be three possibilities:

1. There really is a complete unified theory (or a collection of overlapping formulations), which we will someday discover if we are smart enough.
2. There is no ultimate theory of the universe, just an infinite sequence of theories that describe the universe more and more accurately.
3. There is no theory of the universe: events cannot be predicted beyond a certain extent but occur in a random and arbitrary manner.

Some would argue for the third possibility on the grounds that if there were a complete set of laws, that would infringe God's freedom to change his mind and intervene in the world. It's a bit like the old paradox: can God make a stone so heavy that he can't lift it? But the idea that God might want to change his mind is an example of the fallacy, pointed out by St. Augustine, of imagining God as a being existing in time: time is a property only of the universe that God created. Presumably, he knew what he intended when he set it up!

With the advent of quantum mechanics, we have come to recognize that events cannot be predicted with complete accuracy but that there is always a degree of uncertainty. If one likes, one could ascribe this randomness to the intervention of God, but it would be a very strange kind of intervention: there is no evidence that it is directed toward any purpose. Indeed, if it were, it would by definition not be random. In modern times, we have effectively removed the third possibility above by redefining the goal of science: our aim is to formulate a set of laws that enables us to predict events only up to the limit set by the uncertainty principle.

The second possibility, that there is an infinite sequence of more and more refined theories, is in agreement with all our experience so far. On many occasions we have increased the sensitivity of our measurements or made a new class of observations, only to discover new phenomena that were not predicted by the existing theory, and to account for these we have had to develop a more advanced theory. It would therefore not be very surprising if the present generation of grand unified theories was wrong in claiming that nothing essentially new will happen between the electroweak unification energy of about 100 GeV and the grand unification energy of about a thousand million million GeV. We might indeed expect to find several new layers of structure more

basic than the quarks and electrons that we now regard as “elementary” particles.

However, it seems that gravity may provide a limit to this sequence of “boxes within boxes.” If one had a particle with an energy above what is called the Planck energy, ten million million million GeV (1 followed by nineteen zeros), its mass would be so concentrated that it would cut itself off from the rest of the universe and form a little black hole. Thus it does seem that the sequence of more and more refined theories should have some limit as we go to higher and higher energies, so that there should be some ultimate theory of the universe. Of course, the Planck energy is a very long way from the energies of around a hundred GeV, which are the most that we can produce in the laboratory at the present time. We shall not bridge that gap with particle accelerators in the foreseeable future! The very early stages of the universe, however, are an arena where such energies must have occurred. I think that there is a good chance that the study of the early universe and the requirements of mathematical consistency will lead us to a complete unified theory within the lifetime of some of us who are around today, always presuming we don’t blow ourselves up first.

What would it mean if we actually did discover the ultimate theory of the universe? As was explained in Chapter 1, we could never be quite sure that we had indeed found the correct theory, since theories can’t be proved. But if the theory was mathematically consistent and always gave predictions that agreed with observations, we could be reasonably confident that it was the right one. It would bring to an end a long and glorious chapter in the history of humanity’s intellectual struggle to understand the universe. But it would also revolutionize the ordinary person’s understanding of the laws that govern the universe. In Newton’s time it was possible for an educated person to have a grasp of the whole of human knowledge, at least in outline. But since then, the pace of the development of science has made this impossible. Because theories are always being changed to account for new observations, they are never properly digested or simplified so that ordinary people can understand them. You have to be a specialist, and even then you can only hope to have a proper grasp of a small proportion of the scientific theories. Further, the rate of progress is so rapid that what one learns at school or university is always a bit out of date. Only a few people can keep up with the rapidly advancing frontier of knowledge, and they have to devote their whole time to it and specialize in a small area. The rest of the population has little idea of the advances that are being made or the excitement they are generating. Seventy years ago, if Eddington is to be believed, only two people understood the general theory of relativity. Nowadays tens of thousands of university graduates do, and many millions of people are at least familiar with the idea. If a complete unified theory was discovered, it would only be a matter of time before it was digested and simplified in the same way and taught in schools, at least in outline. We would then all be able to have some understanding of the laws that govern the universe and are responsible for our existence.

Even if we do discover a complete unified theory, it would not mean that we would be able to predict events in general, for two reasons. The first is the limitation that the uncertainty principle of quantum mechanics sets on our powers of prediction. There is nothing we can do to get around that. In practice, however, this first limitation is less restrictive than the second one. It arises from the fact that we could not solve the equations of the theory exactly, except in very simple situations. (We cannot even solve exactly for the motion of three bodies in Newton’s theory of gravity, and the difficulty increases with the number of bodies and the complexity of the theory.) We already know the laws that govern the behavior of matter under all but the most extreme conditions. In particular, we know the basic laws that underlie all of chemistry and biology. Yet we have certainly not reduced these subjects to the status of solved problems: we have, as yet, had little success in predicting human behavior from mathematical equations! So even if we do find a complete set of basic laws, there will still be in the years ahead the intellectually challenging task of developing better approximation methods, so that we can make useful predictions of the probable outcomes in complicated and realistic situations. A complete, consistent, unified theory is only the first step: our goal is a complete *understanding* of the events around us, and of our own existence.

[PREVIOUS](#)
[NEXT](#)

