2

SPACE AND TIME

Our present ideas about the motion of bodies date back to Galileo and Newton. Before them people believed Aristotle, who said that the natural state of a body was to be at rest and that it moved only if driven by a force or impulse. It followed that a heavy body should fall faster than a light one, because it would have a greater pull toward the earth.

The Aristotelian tradition also held that one could work out all the laws that govern the universe by pure thought: it was not necessary to check by observation. So no one until Galileo bothered to see whether bodies of different weight did in fact fall at different speeds. It is said that Galileo demonstrated that Aristotle's belief was false by dropping weights from the leaning tower of Pisa. The story is almost certainly untrue, but Galileo did do something equivalent: he rolled balls of different weights down a smooth slope. The situation is similar to that of heavy bodies falling vertically, but it is easier to observe because the speeds are smaller. Galileo's measurements indicated that each body increased its speed at the same rate, no matter what its weight. For example, if you let go of a ball on a slope that drops by one meter for every ten meters you go along, the ball will be traveling down the slope at a speed of about one meter per second after one second, two meters per second after two seconds, and so on, however heavy the ball. Of course a lead weight would fall faster than a feather, but that is only because a feather is slowed down by air resistance. If one drops two bodies that don't have much air resistance, such as two different lead weights, they fall at the same rate.

Galileo's measurements were used by Newton as the basis of his laws of motion. In Galileo's experiments, as a body rolled down the slope it was always acted on by the same force (its weight), and the effect was to make it constantly speed up. This showed that the real effect of a force is always to change the speed of a body, rather than just to set it moving, as was previously thought. It also meant that whenever a body is not acted on by any force, it will keep on moving in a straight line at the same speed. This idea was first stated explicitly in Newton's *Principia Mathematica,* published in 1687, and is known as Newton's first law. What happens to a body when a force does act on it is given by Newton's second law. This states that the body will accelerate, or change its speed, at a rate that is proportional to the force. (For example, the acceleration is twice as great if the force is twice as great.) The acceleration is also smaller the greater the mass (or quantity of matter) of the body. (The same force acting on a body of twice the mass will produce half the acceleration.) A familiar example is provided by a car: the more powerful the engine, the greater the acceleration, but the heavier the car, the smaller the acceleration for the same engine.

In addition to his laws of motion, Newton discovered a law to describe the force of gravity, which states that every body attracts every other body with a force that is proportional to the mass of each body. Thus the force between two bodies would be twice as strong if one of the bodies (say, body *A*) had its mass doubled. This is what you might expect because one could think of the new body *A* as being made of two bodies with the original mass. Each would attract body *B* with the original force. Thus the total force between *A* and *B* would be twice the original force. And if, say, one of the bodies had twice the mass, and the other had three times the mass,

then the force would be six times as strong. One can now see why all bodies fall at the same rate: a body of twice the weight will have twice the force of gravity pulling it down, but it will also have twice the mass. According to Newton's second law, these two effects will exactly cancel each other, so the acceleration will be the same in all cases.

Newton's law of gravity also tells us that the farther apart the bodies, the smaller the force. Newton's law of gravity says that the gravitational attraction of a star is exactly one quarter that of a similar star at half the distance. This law predicts the orbits of the earth, the moon, and the planets with great accuracy. If the law were that the gravitational attraction of a star went down faster with distance, the orbits of the planets would not be elliptical, they would spiral in to the sun. If it went down slower, the gravitational forces from the distant stars would dominate over that from the earth.

The big difference between the ideas of Aristotle and those of Galileo and Newton is that Aristotle believed in a preferred state of rest, which any body would take up if it were not driven by some force or impulse. In particular, he thought that the earth was at rest. But it follows from Newton's laws that there is no unique standard of rest. One could equally well say that body *A* was at rest and body *B* was moving at constant speed with respect to body *A*, or that body *B* was at rest and body *A* was moving. For example, if one sets aside for a moment the rotation of the earth and its orbit round the sun, one could say that the earth was at rest and that a train on it was traveling north at ninety miles per hour or that the train was at rest and the earth was moving south at ninety miles per hour. If one carried out experiments with moving bodies on the train, all Newton's laws would still hold. For instance, playing Ping-Pong on the train, one would find that the ball obeyed Newton's laws just like a ball on a table by the track. So there is no way to tell whether it is the train or the earth that is moving.

The lack of an absolute standard of rest meant that one could not determine whether two events that took place at different times occurred in the same position in space. For example, suppose our

Ping-Pong ball on the train bounces straight up and down, hitting the table twice on the same spot one second apart. To someone on the track, the two bounces would seem to take place about a hundred meters apart, because the train would have traveled that far down the track between the bounces. The nonexistence of absolute rest therefore meant that one could not give an event an absolute position in space, as Aristotle had believed. The positions of events and the distances between them would be different for a person on the train and one on the track, and there would be no reason to prefer one person's position to the other's.

Newton was very worried by this lack of absolute position, or absolute space, as it was called, because it did not accord with his idea of an absolute God. In fact, he refused to accept absolute space, even though it was implied by his laws. He was severely criticized for this irrational belief by many people, most notably by Bishop Berkeley, a philosopher who believed that all material objects and space and time are an illusion. When the famous Dr. Johnson was told of Berkeley's opinion, he cried, "I refute it thus!" and stubbed his toe on a large stone.

Both Aristotle and Newton believed in absolute time. That is, they believed that one could unambiguously measure the interval of time between two events, and that this time would be the same whoever measured it, provided they used a good clock. Time was completely separate from and independent of space. This is what most people would take to be the commonsense view. However, we have had to change our ideas about space and time. Although our apparently commonsense notions work well when dealing with things like apples, or planets that travel comparatively slowly, they don't work at all for things moving at or near the speed of light.

The fact that light travels at a finite, but very high, speed was first discovered in 1676 by the Danish astronomer Ole Christensen Roemer. He observed that the times at which the moons of Jupiter appeared to pass behind Jupiter were not evenly spaced, as one would expect if the moons went round Jupiter at a constant rate. As the earth and Jupiter orbit around the sun, the distance between them varies. Roemer noticed that eclipses of Jupiter's moons appeared later the farther we were from Jupiter. He argued that this was because the light from the moons took longer to reach us when we were farther away. His measurements of the variations in the distance of the earth from Jupiter were, however, not very accurate, and so his value for the speed of light was 140,000 miles per second, compared to the modern value of 186,000 miles per second. Nevertheless, Roemer's achievement, in not only proving that light travels at a finite speed, but also in measuring that speed, was remarkable—coming as it did eleven years before Newton's publication of *Principia Mathematica*.

A proper theory of the propagation of light didn't come until 1865 when the British physicist James Clerk Maxwell succeeded in unifying the partial theories that up to then had been used to describe the forces of electricity and magnetism. Maxwell's equations predicted that there could be wavelike disturbances in the combined electromagnetic field, and that these would travel at a fixed speed, like ripples on a pond. If the wavelength of these waves (the distance between one wave crest and the next) is a meter or more, they are what we now call radio waves. Shorter wavelengths are known as microwaves (a few centimeters) or infrared (more than a ten thousandth of a centimeter). Visible light has a wavelength of between only forty and eighty millionths of a centimeter. Even shorter wavelengths are known as ultraviolet, X rays, and gamma rays.

Maxwell's theory predicted that radio or light waves should travel at a certain fixed speed. But Newton's theory had got rid of the idea of absolute rest, so if light was supposed to travel at a fixed speed, one would have to say what that fixed speed was to be measured relative to. It was therefore suggested that there was a substance called the "ether" that was present everywhere, even in "empty" space. Light waves should travel through the ether as sound waves travel through air, and their speed should therefore be relative to the ether. Different observers, moving relative to the ether, would see light coming toward them at different speeds, but light's speed relative to the ether would remain fixed. In particular, as the earth was moving through the ether on its orbit round the sun, the speed of light measured in the direction of the earth's motion through the ether (when we were moving toward the source of the light) should be higher than the speed of light at right angles to that motion (when we are not moving toward the source). In 1887 Albert Michelson (who later became the first American to receive the Nobel prize for physics) and Edward Morley carried out a very careful experiment at the Case School of Applied Science in Cleveland. They compared the speed of light in the direction of the earth's motion with that at right angles to the earth's motion. To their great surprise, they found they were exactly the same!

Between 1887 and 1905 there were several attempts, most notably by the Dutch physicist Hendrik Lorentz, to explain the result of the Michelson–Morley experiment in terms of objects contracting and clocks slowing down when they moved through the ether. However, in a famous paper in 1905, a hitherto unknown clerk in the Swiss patent office, Albert Einstein, pointed out that the whole idea of an ether was unnecessary, providing one was willing to abandon the idea of absolute time. A similar point was made a few weeks later by a leading French mathematician, Henri Poincaré. Einstein's arguments were closer to physics than those of Poincaré, who regarded this problem as mathematical. Einstein is usually given the credit for the new theory, but Poincaré is remembered by having his name attached to an important part of it.

The fundamental postulate of the theory of relativity, as it was called, was that the laws of science should be the same for all freely moving observers, no matter what their speed. This was true for Newton's laws of motion, but now the idea was extended to include Maxwell's theory and the speed of light: all observers should measure the same speed of light, no matter how fast they are moving. This simple idea has some remarkable consequences. Perhaps the best known are the equivalence of mass and energy, summed up in Einstein's famous equation $E=mc^2$ (where E is energy, m is mass and c is the speed of light), and the law that nothing may travel faster than the speed of light. Because of the equivalence of energy and mass, the energy which an object has due to its motion will add to its mass. In other words, it will make it harder to increase its speed. This effect is only really significant

for objects moving at speeds close to the speed of light. For example, at 10 percent of the speed of light an object's mass is only 0.5 percent more than normal, while at 90 percent of the speed of light it would be more than twice its normal mass. As an object approaches the speed of light, its mass rises ever more quickly, so it takes more and more energy to speed it up further. It can in fact never reach the speed of light, because by then its mass would have become infinite, and by the equivalence of mass and energy, it would have taken an infinite amount of energy to get it there. For this reason, any normal object is forever confined by relativity to move at speeds slower than the speed of light. Only light, or other waves that have no intrinsic mass, can move at the speed of light.

An equally remarkable consequence of relativity is the way it has revolutionized our ideas of space and time. In Newton's theory, if a pulse of light is sent from one place to another, different observers would agree on the time that the journey took (since time is absolute), but will not always agree on how far the light traveled (since space is not absolute). Since the speed of the light is just the distance it has traveled divided by the time it has taken, different observers would measure different speeds for the light. In relativity, on the other hand, all observers *must* agree on how fast light travels. They still, however, do not agree on the distance the light has traveled, so they must therefore now also disagree over the time it has taken. (The time it has taken is, after all, just the light's speed—which the observers agree on—multiplied by the distance the light has traveled—which they do not agree on.) In other words, the theory of relativity put an end to the idea of absolute time! It appeared that each observer must have his own measure of time, as recorded by a clock carried with him, and that identical clocks carried by different observers would not necessarily agree.

Each observer could use radar to say where and when an event took place by sending out a pulse of light or radio waves. Part of the pulse is reflected back at the event and the observer measures the time at which he receives the echo. The time of the event is then said to be the time halfway between when the pulse was sent and the time when the reflection was received back; the distance of the event is half the time taken for this round trip, multiplied by the

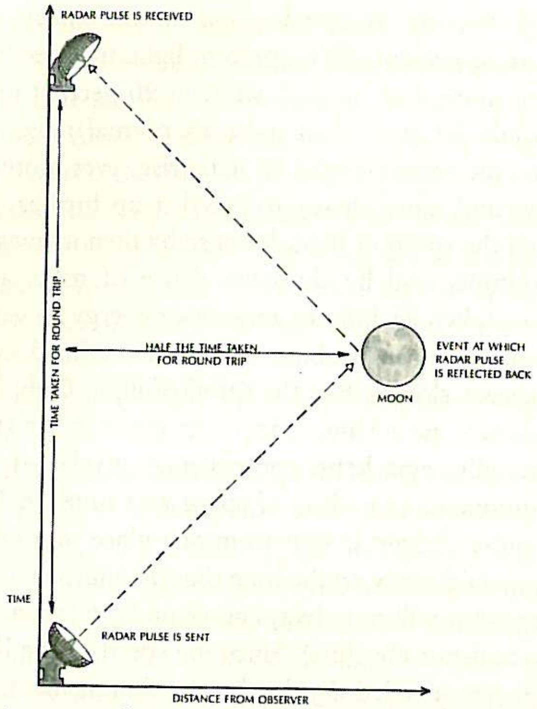

FIGURE 2-1 Time is measured vertically, and the distance from the observer is measured horizontally. The observer's path through space and time is shown as the vertical line on the left. The paths of light rays to and from the event are the diagonal lines.

speed of light. (An event, in this sense, is something that takes place at a single point in space, at a specified point in time.) This idea is shown in Fig. 2.1, which is an example of a space-time diagram. Using this procedure, observers who are moving relative to each other will assign different times and positions to the same event. No particular observer's measurements are any more correct than any other observer's, but all the measurements are related. Any observer can work out precisely what time and position any other observer will assign to an event, provided he knows the other observer's relative velocity.

Nowadays we use just this method to measure distances precisely, because we can measure time more accurately than length. In effect, the meter is defined to be the distance traveled by light in 0.000000003335640952 seconds, as measured by a cesium clock.

(The reason for that particular number is that it corresponds to the historical definition of the meter—in terms of two marks on a particular platinum bar kept in Paris.) Equally, we can use a more convenient, new unit of length called a light-second. This is simply defined as the distance that light travels in one second. In the theory of relativity, we now define distance in terms of time and the speed of light, so it follows automatically that every observer will measure light to have the same speed (by definition, 1 meter per 0.000000003335640952 seconds). There is no need to introduce the idea of an ether, whose presence anyway cannot be detected, as the Michelson–Morley experiment showed. The theory of relativity does, however, force us to change fundamentally our ideas of space and time. We must accept that time is not completely separate from and independent of space, but is combined with it to form an object called space-time.

It is a matter of common experience that one can describe the position of a point in space by three numbers, or coordinates. For instance, one can say that a point in a room is seven feet from one wall, three feet from another, and five feet above the floor. Or one could specify that a point was at a certain latitude and longitude and a certain height above sea level. One is free to use any three suitable coordinates, although they have only a limited range of validity. One would not specify the position of the moon in terms of miles north and miles west of Piccadilly Circus and feet above sea level. Instead, one might describe it in terms of distance from the sun, distance from the plane of the orbits of the planets, and the angle between the line joining the moon to the sun and the line joining the sun to a nearby star such as Alpha Centauri. Even these coordinates would not be of much use in describing the position of the sun in our galaxy or the position of our galaxy in the local group of galaxies. In fact, one may describe the whole universe in terms of a collection of overlapping patches. In each patch, one can use a different set of three coordinates to specify the position of a point.

An event is something that happens at a particular point in space and at a particular time. So one can specify it by four numbers or coordinates. Again, the choice of coordinates is arbitrary; one can use any three well-defined spatial coordinates and any measure of time. In relativity, there is no real distinction between the space and time coordinates, just as there is no real difference between any two space coordinates. One could choose a new set of coordinates in which, say, the first space coordinate was a combination of the old first and second space coordinates. For instance, instead of measuring the position of a point on the earth in miles north of Piccadilly and miles west of Piccadilly, one could use miles northeast of Piccadilly, and miles northwest of Piccadilly. Similarly, in relativity, one could use a new time coordinate that was the old time (in seconds) plus the distance (in light-seconds) north of Piccadilly.

It is often helpful to think of the four coordinates of an event as specifying its position in a four-dimensional space called space-time. It is impossible to imagine a four-dimensional space. I personally find it hard enough to visualize three-dimensional space! However, it is easy to draw diagrams of two-dimensional spaces, such as the surface of the earth. (The surface of the earth is two-dimensional because the position of a point can be specified by two coordinates, latitude and longtitude.) I shall generally use diagrams in which time increases upward and one of the spatial dimensions is shown horizontally. The other two spatial dimensions are ignored or, sometimes, one of them is indicated by perspective. (These are called space-time diagrams, like Fig. 2.1.) For example, in Fig. 2.2 time is measured upwards in years and the distance along the line from the sun to Alpha Centauri is measured horizontally in miles. The paths of the sun and of Alpha Centauri through space-time are shown as the vertical lines on the left and right of the diagram. A ray of light from the sun follows the diagonal line, and takes four years to get from the sun to Alpha Centauri.

As we have seen, Maxwell's equations predicted that the speed of light should be the same whatever the speed of the source, and this has been confirmed by accurate measurements. It follows from this that if a pulse of light is emitted at a particular time at a particular point in space, then as time goes on it will spread out as a sphere of light whose size and position are independent of the speed of the source. After one millionth of a second the light will have spread out to form a sphere with a radius of 300 meters; after

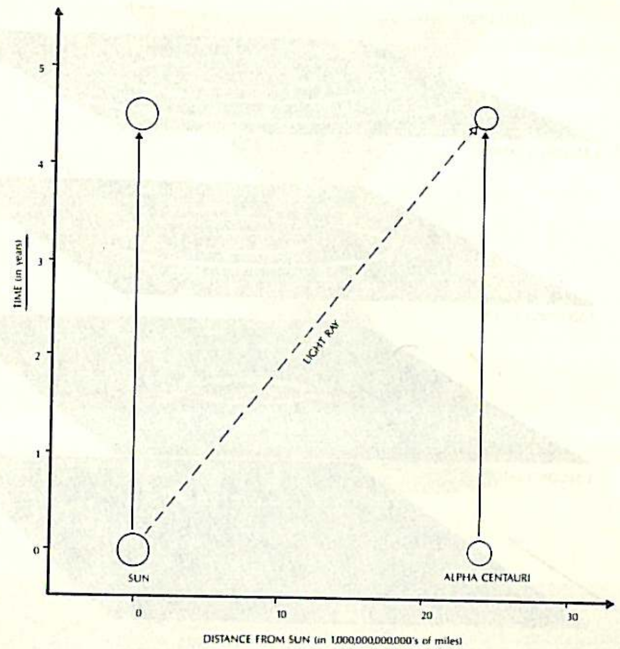

FIGURE 2.2

two millionths of a second, the radius will be 600 meters; and so on. It will be like the ripples that spread out on the surface of a pond when a stone is thrown in. The ripples spread out as a circle that gets bigger as time goes on. If one thinks of a three-dimensional model consisting of the two-dimensional surface of the pond and the one dimension of time, the expanding circle of ripples will mark out a cone whose tip is at the place and time at which the stone hit the water (Fig. 2.3). Similarly, the light spreading out from an event forms a three-dimensional cone in the four-dimensional space-time. This cone is called the future light cone of the event. In the same way we can draw another cone, called the past light cone, which is the set of events from which a pulse of light is able to reach the given event (Fig. 2.4).

The past and the future light cones of an event *P* divide space-time into three regions (Fig. 2.5). The absolute future of the event is the region inside the future light cone of *P*. It is the set of all events that can possibly be affected by what happens at *P*.

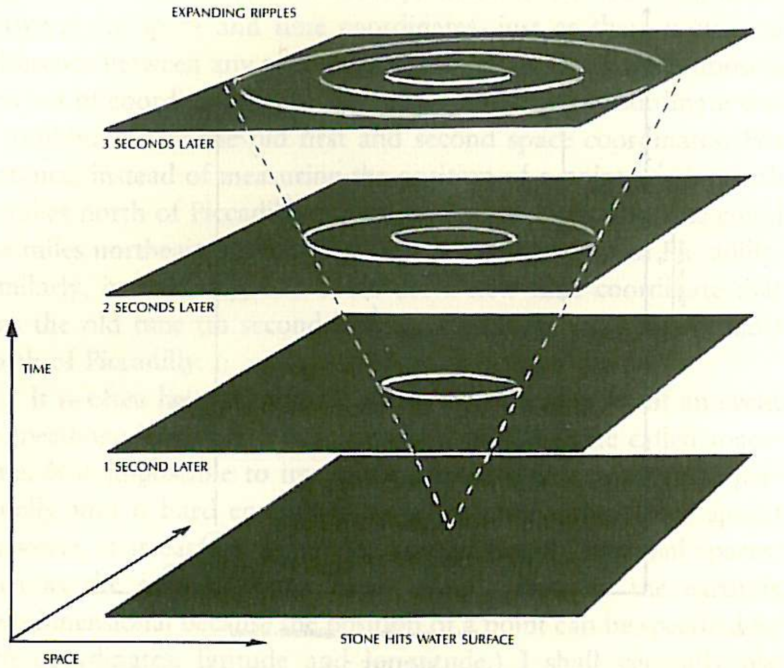

FIGURE 2.3

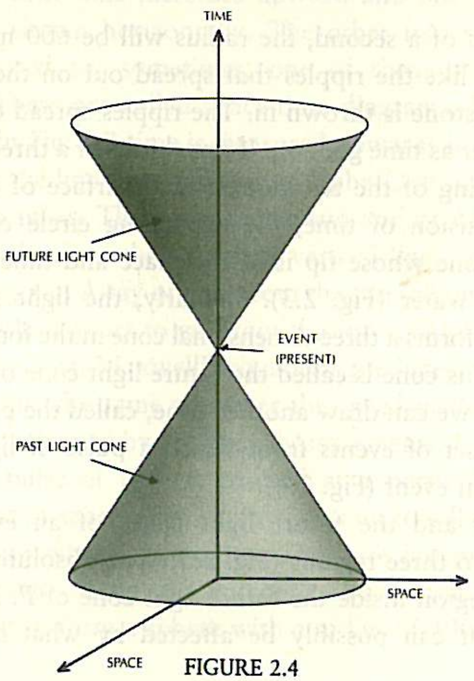

FIGURE 2.4

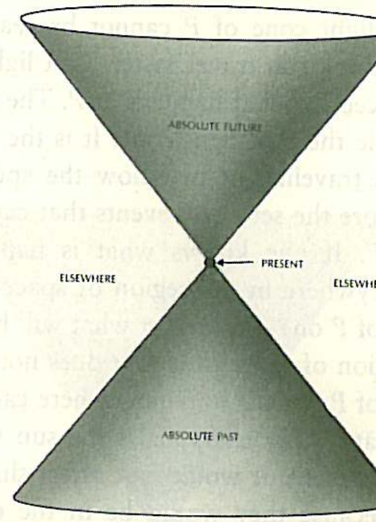

FIGURE 2.5

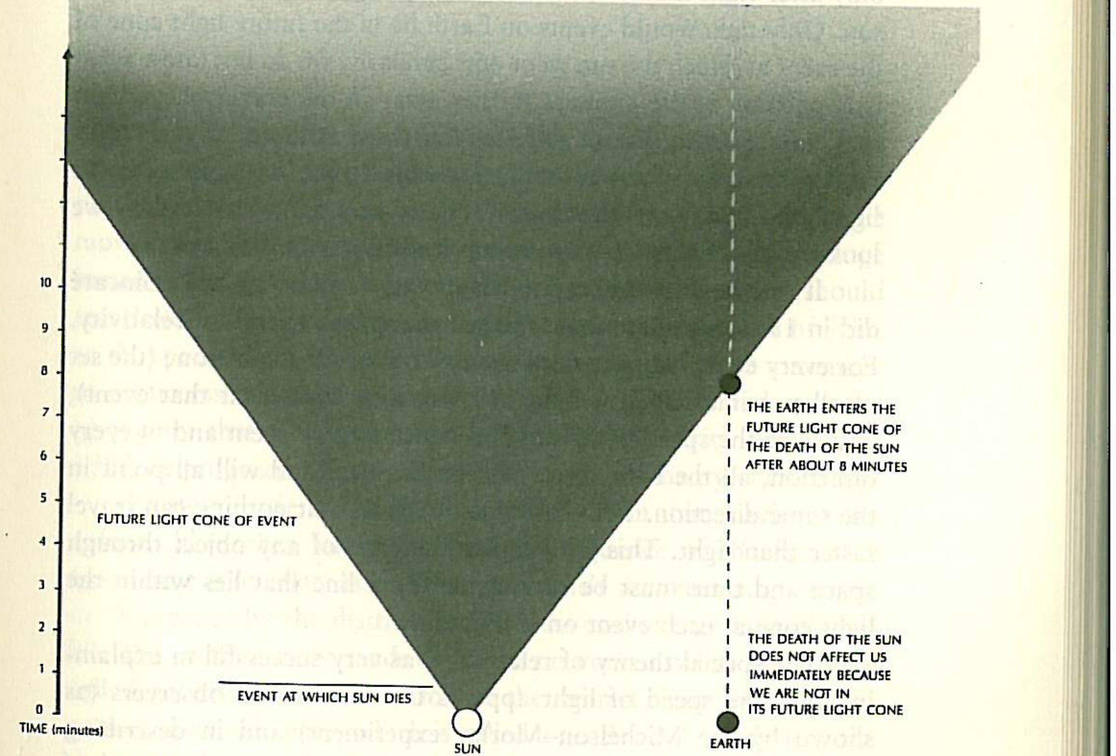

FIGURE 2.6

Events outside the light cone of *P* cannot be reached by signals from *P* because nothing can travel faster than light. They cannot therefore be influenced by what happens at *P*. The absolute past of *P* is the region inside the past light cone. It is the set of all events from which signals traveling at or below the speed of light can reach *P*. It is therefore the set of all events that can possibly affect what happens at *P*. If one knows what is happening at some particular time everywhere in the region of space that lies within the past light cone of *P* one can predict what will happen at *P*. The elsewhere is the region of space-time that does not lie in the future or past light cones of *P*. Events in the elsewhere cannot affect or be affected by events at *P*. For example, if the sun were to cease to shine at this very moment, it would not affect things on Earth at the present time because they would be in the elsewhere of the event when the sun went out (Fig. 2.6). We would know about it only after eight minutes, the time it takes light to reach us from the sun. Only then would events on Earth lie in the future light cone of the event at which the sun went out. Similarly, we do not know what is happening at the moment farther away in the universe: the light that we see from distant galaxies left them millions of years ago, and in the case of the most distant object that we have seen, the light left some eight thousand million years ago. Thus, when we look at the universe, we are seeing it as it was in the past.

If one neglects gravitational effects, as Einstein and Poincaré did in 1905, one has what is called the special theory of relativity. For every event in space-time we may construct a light cone (the set of all possible paths of light in space-time emitted at that event), and since the speed of light is the same at every event and in every direction, all the light cones will be identical and will all point in the same direction. The theory also tells us that nothing can travel faster than light. This means that the path of any object through space and time must be represented by a line that lies within the light cone at each event on it (Fig. 2.7).

The special theory of relativity was very successful in explaining that the speed of light appears the same to all observers (as shown by the Michelson–Morley experiment) and in describing what happens when things move at speeds close to the speed of

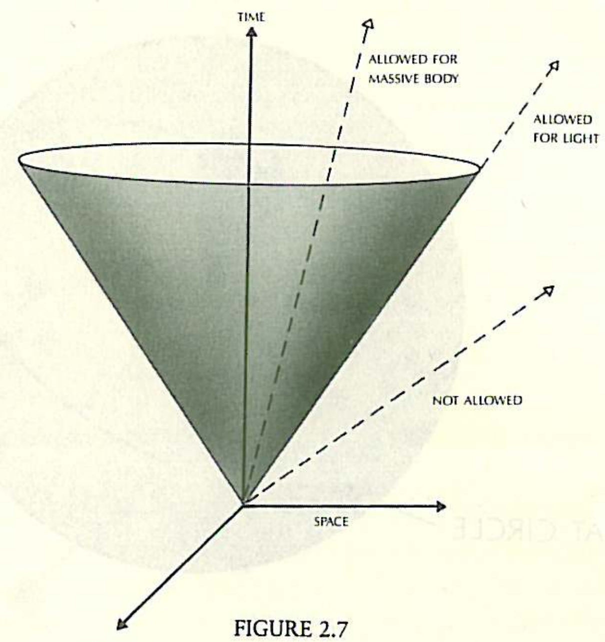

FIGURE 2.7

light. However, it was inconsistent with the Newtonian theory of gravity, which said that objects attracted each other with a force that depended on the distance between them. This meant that if one moved one of the objects, the force on the other one would change instantaneously. Or in other words, gravitational effects should travel with infinite velocity, instead of at or below the speed of light, as the special theory of relativity required. Einstein made a number of unsuccessful attempts between 1908 and 1914 to find a theory of gravity that was consistent with special relativity. Finally, in 1915, he proposed what we now call the general theory of relativity.

Einstein made the revolutionary suggestion that gravity is not a force like other forces, but is a consequence of the fact that space-time is not flat, as had been previously assumed: it is curved, or "warped," by the distribution of mass and energy in it. Bodies like the earth are not made to move on curved orbits by a force called gravity; instead, they follow the nearest thing to a straight path in a curved space, which is called a geodesic. A geodesic is the shortest (or longest) path between two nearby points. For example,

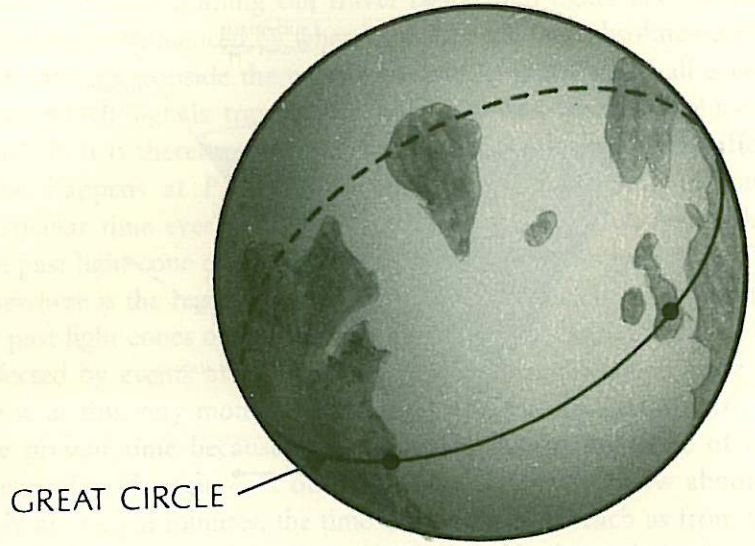

FIGURE 2.8

the surface of the earth is a two-dimensional curved space. A geodesic on the earth is called a great circle, and is the shortest route between two points (Fig. 2.8). As the geodesic is the shortest path between any two airports, this is the route an airline navigator will tell the pilot to fly along. In general relativity, bodies always follow straight lines in four-dimensional space-time, but they nevertheless appear to us to move along curved paths in our three-dimensional space. (This is rather like watching an airplane flying over hilly ground. Although it follows a straight line in three-dimensional space, its shadow follows a curved path on the two-dimensional ground.)

The mass of the sun curves space-time in such a way that although the earth follows a straight path in four-dimensional space-time, it appears to us to move along a circular orbit in three-dimensional space. In fact, the orbits of the planets predicted by general relativity are almost exactly the same as those predicted by the Newtonian theory of gravity. However, in the case of Mercury, which, being the nearest planet to the sun, feels the

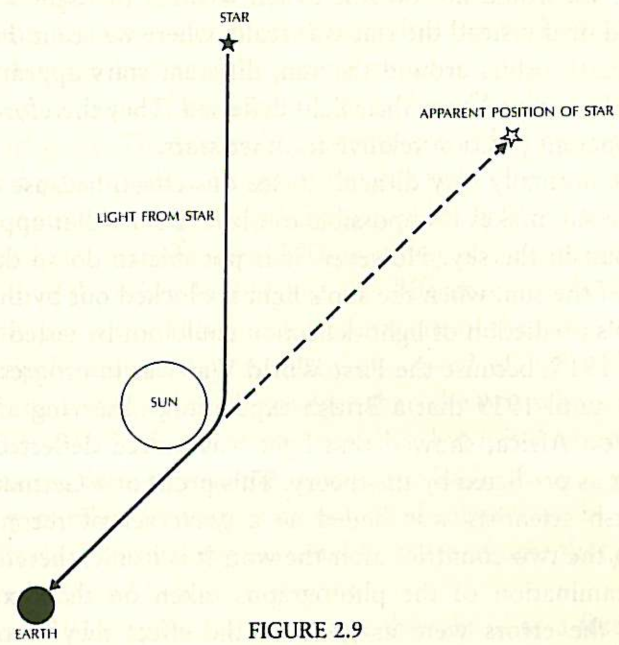

FIGURE 2.9

strongest gravitational effects, and has a rather elongated orbit, general relativity predicts that the long axis of the ellipse should rotate about the sun at a rate of about one degree in ten thousand years. Small though this effect is, it had been noticed before 1915 and served as one of the first confirmations of Einstein's theory. In recent years the even smaller deviations of the orbits of the other planets from the Newtonian predictions have been measured by radar and found to agree with the predictions of general relativity.

Light rays too must follow geodesics in space-time. Again, the fact that space is curved means that light no longer appears to travel in straight lines in space. So general relativity predicts that light should be bent by gravitational fields. For example, the theory predicts that the light cones of points near the sun would be slightly bent inward, on account of the mass of the sun. This means that light from a distant star that happened to pass near the sun would be deflected through a small angle, causing the star to appear in a different position to an observer on the earth (Fig. 2.9). Of course, if the light from the star always passed close to the sun, we would not be able to tell whether the light was being deflected or if instead the star was really where we see it. However, as the earth orbits around the sun, different stars appear to pass behind the sun and have their light deflected. They therefore change their apparent position relative to other stars.

It is normally very difficult to see this effect, because the light from the sun makes it impossible to observe stars that appear near to the sun in the sky. However, it is possible to do so during an eclipse of the sun, when the sun's light is blocked out by the moon. Einstein's prediction of light deflection could not be tested immediately in 1915, because the First World War was in progress, and it was not until 1919 that a British expedition, observing an eclipse from West Africa, showed that light was indeed deflected by the sun, just as predicted by the theory. This proof of a German theory by British scientists was hailed as a great act of reconciliation between the two countries after the war. It is ironic, therefore, that later examination of the photographs taken on that expedition showed the errors were as great as the effect they were trying to measure. Their measurement had been sheer luck, or a case of knowing the result they wanted to get, not an uncommon occurrence in science. The light deflection has, however, been accurately confirmed by a number of later observations.

Another prediction of general relativity is that time should appear to run slower near a massive body like the earth. This is because there is a relation between the energy of light and its frequency (that is, the number of waves of light per second): the greater the energy, the higher the frequency. As light travels upward in the earth's gravitational field, it loses energy, and so its frequency goes down. (This means that the length of time between one wave crest and the next goes up.) To someone high up, it would appear that everything down below was taking longer to happen. This prediction was tested in 1962, using a pair of very accurate clocks mounted at the top and bottom of a water tower. The clock at the bottom, which was nearer the earth, was found to run slower, in exact agreement with general relativity. The difference in the speed of clocks at different heights above the earth is now of considerable practical importance, with the advent of very

accurate navigation systems based on signals from satellites. If one ignored the predictions of general relativity, the position that one calculated would be wrong by several miles!

Newton's laws of motion put an end to the idea of absolute position in space. The theory of relativity gets rid of absolute time. Consider a pair of twins. Suppose that one twin goes to live on the top of a mountain while the other stays at sea level. The first twin would age faster than the second. Thus, if they met again, one would be older than the other. In this case, the difference in ages would be very small, but it would be much larger if one of the twins went for a long trip in a spaceship at nearly the speed of light. When he returned, he would be much younger than the one who stayed on Earth. This is known as the twins paradox, but it is a paradox only if one has the idea of absolute time at the back of one's mind. In the theory of relativity there is no unique absolute time, but instead each individual has his own personal measure of time that depends on where he is and how he is moving.

Before 1915, space and time were thought of as a fixed arena in which events took place, but which was not affected by what happened in it. This was true even of the special theory of relativity. Bodies moved, forces attracted and repelled, but time and space simply continued, unaffected. It was natural to think that space and time went on forever.

The situation, however, is quite different in the general theory of relativity. Space and time are now dynamic quantities: when a body moves, or a force acts, it affects the curvature of space and time—and in turn the structure of space-time affects the way in which bodies move and forces act. Space and time not only affect but also are affected by everything that happens in the universe. Just as one cannot talk about events in the universe without the notions of space and time, so in general relativity it became meaningless to talk about space and time outside the limits of the universe.

In the following decades this new understanding of space and time was to revolutionize our view of the universe. The old idea of an essentially unchanging universe that could have existed, and could continue to exist, forever was replaced by the notion of a dynamic, expanding universe that seemed to have begun a finite time ago, and that might end at a finite time in the future. That revolution forms the subject of the next chapter. And years later, it was also to be the starting point for my work in theoretical physics. Roger Penrose and I showed that Einstein's general theory of relativity implied that the universe must have a beginning and, possibly, an end.