to the philosophy of space and time in the context of these new physical theories.

From Space and Time to Spacetime

The Origins of the Special Theory of Relativity

We have seen that whereas Newton posited "space itself" as the reference object relative to which accelerations generated observable inertial forces, uniform motion with respect to space itself was deemed to have no observable consequences. This followed from Galileo's famous observation that in an enclosed laboratory, one could not tell which state of uniform motion the laboratory was in by performing any mechanical experiment. But it remained conceivable that some other, nonmechanical, phenomena would depend upon the uniform motion of the apparatus with respect to space itself in some way. This motion would then reveal itself in an observational consequence.

In the nineteenth century, hope for this came out of the reduction of light to electromagnetic radiation. According to J. C. Maxwell's theory of electricity and magnetism, electromagnetic waves, of which light waves are a species, are predicted to have a definite velocity with respect to an observer. This velocity should be the same in all directions and independent of the velocity of the source of the light with respect to the observer. An observer at rest in a tank of water will determine a speed of sound in the water that is the same in each direction. This speed of sound will be completely independent of the motion of the source of the sound in the water. Once the water wave is generated, its speed depends only on the properties of the water in which the wave is traveling. So it should be for light, with the medium of transmission of light (the stuff that is for light as water is for the sound) called the "aether."

An observer who himself moves through the water in the tank will not see the speed of sound the same in all directions, as he will be catching up to the sound in one direction and running away from it in the opposite direction. So an observer in motion with respect to the aether should be able to detect this motion, even if it is uniform, unaccelerated motion, by measuring the velocity of light in all directions. If we assume that an observer at rest in the aether will be at rest in one of the inertial frames of mechanics in which no mechanical, inertial forces are generated, it becomes plausible to identify the aether with Newton's space itself. The assumption was always made in the nineteenth century, and in a reinterpreted version, it remains correct in relativity theory. We could then use experiments with light to determine our uniform motion with respect to space itself.

An ingenious series of experiments was designed to detect which state of uniform motion was the state at rest in the aether or in space itself. These worked by sending light out from a point along different paths and then bringing the light back to its point of origin. The light should take different amounts of time to traverse the different paths, depending on the length of the paths and on the state of motion of the apparatus in the aether. Changing the orientation of the apparatus, or letting the motion of the earth do that for us as the earth rotated on its axis and traveled in its orbit around the sun, would change the relative times taken by the light to travel the different paths. Such a change in times could be detected at the origin of the light by an observer, who would see a shift in the position of the interference lines, alternating lines of light and darkness that are generated when the two returning beams of light meet and have regions of varying intensity add or subtract from one another. (See Figure 2.1.)

When the experiments were performed, much to the astonishment of those performing them, no detectable difference in travel times for the light could be discerned. It was as if the light traveled with the same fixed speed, the speed predicted for light by the theory in the rest frame in the aether, in every laboratory frame that was in uniform motion. (These "null results" don't hold, incidentally, when the apparatus is in nonuniform motion. Rotation can be detected, for example, by a ring laser gyro, which detects the change in speed of light in opposite directions around a circular path as the laboratory rotates.) Now it might seem that this surprising null result might be due to some peculiarity of light or electromagnetism. If one thinks about why the speed of the signal should vary when the laboratory in motion with respect to the medium of transmission of the signal, however, one quickly sees that a very fundamental intuition about motion is being challenged here. That intuition is that, for example, if we run after a moving thing, it will be moving more slowly with respect to us than it will be to someone who didn't join in the chase.

One could try to explain away the surprising results in a number of ways. One suggestion was that the earth in its motion dragged the local

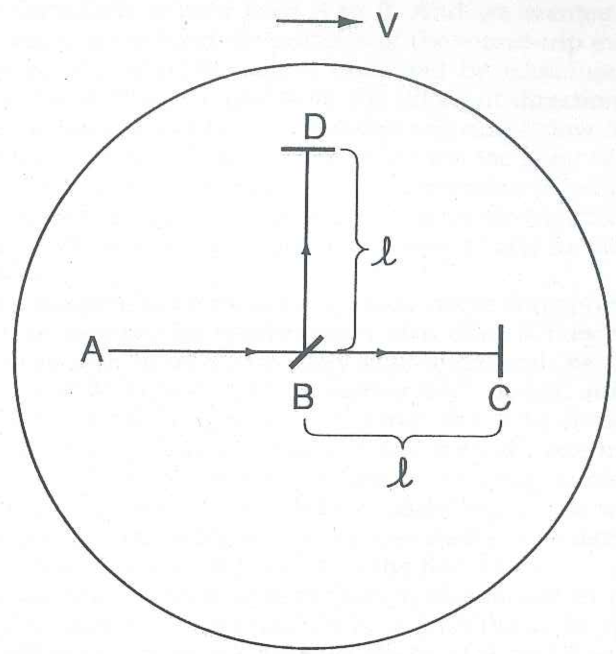

Figure 2.1 The Michelson-Morley experiment. A beam of light is split into two beams at the half-silvered mirror *B*. One beam goes to mirror *C* and is reflected, the other to mirror *D*. If the apparatus is moving through the aether, the medium of transmission of light hypothesized by the older wave theory, in the direction shown by arrow v, the light should take longer to travel path *BCB* of length l than it does to travel path *BDB*, also of length l. If the apparatus is then rotated 90 degrees, the difference in time along the paths is reversed. But no such change is detected when the experiment is carried out. This remains true even if the path length *BC* is made different from length *BD*. In general no round-trip experiment reveals motion through the aether of the laboratory.

aether with it, so that the portion of aether near the earth was always at rest with respect to the earth and the apparatus. Such a claim would, however, result in conflict with well-established astronomical observations.

A series of compensatory theories were invented to explain the unexpected null results. If we assume that the apparatus shrinks in its length in its direction of motion with respect to the aether, and further assume that all physical processes measured by apparatus clocks slow down when these clocks are put in motion with respect to the aether, one could then explain away as appearance the seeming sameness of speed of light in all directions. Although the light actually was moving at different speeds with respect to the apparatus in different directions, the observational consequences expected from this were exactly canceled out by the changes induced (by the motion of the apparatus through the aether) on the components of the apparatus that one used to determine speeds—lengths and time intervals as measured by rods and clocks. The net result would be, then, once more to make uniform motion with respect to space undetectable by any experimental means!

It was Einstein's brilliant suggestion to take the appearance that light has the same velocity in all directions in every uniformly moving state of motion as indicative of reality. Why not posit, he argued, that what appears to be the case from the round-trip experiments really is the case? For each uniformly moving observer, light in vacuum travels at the speed predicted by the theory of electromagnetism in every direction. It is important to note just how radical a proposal this is. If a light beam is moving away from an observer in a given direction at speed c, and a second observer is traveling in the direction of the propagated light with, say, speed v, with respect to the first observer, we take it to be the case that the light is traveling with speed c, and not with speed $c - v$ as intuition tells us, with respect to the second observer as well.

How could this be? The core of Einstein's argument is an insightful critique of the notion of simultaneity for events at a distance from one another. What does it mean for two happenings at a spatial distance from each other to occur at the same time? In pre-Einsteinian thought, we just assume that if two events occur at the same time for one observer, they will occur at the same time for all observers. It is a challenge to this last notion that provides the main difference between space and time as earlier understood and spacetime as understood in Einstein's so-called special theory of relativity.

Einstein argues that if we are to determine the speed of light in a given direction, we might think to get around the null results of the round-trip experiments by directly measuring the speed of light from one point, A, to another, B. But we could only do this if we could determine the distance between the points and the time taken by the light to get from A to B, speed being the distance divided by the time. But to get the time interval between emission and reception of a light signal requires that we be able to synchronize clocks at the two points so that they read "zero" at the same moment. How could this synchronization be done?

If we could transport a clock instantaneously from A to B, we could establish synchronization by synchronizing two clocks at A and instantly shifting one to B. But, Einstein assumes, objects cannot be transported from one place to another in no time at all. He assumes, in fact, that the speed of light in a vacuum is a limiting speed faster than which nothing can travel. Well then, why not synchronize two clocks at A, move one at some speed or other to B, and assume that when a clock at A reads value n and the clock at B reads n the two events are simultaneous?

At this point we must remember the point of trying to establish simultaneity for distant events. We wanted to do this so that we could

determine the speed of light from A to B. And we wanted to do that so that we could get around the problem of the round-trip experiments' giving null results, a phenomenon explained by combining the idea that the light had different speeds in the different directions with the compensatory claims about how rods shrink and clocks slow down when moving with respect to the aether. Remember that the point of the round-trip experiments in the first place was to determine in which state of motion the speed of light was really the same in all directions in order to determine which state of motion really was at rest in the aether or in space itself.

But if the compensatory theory is correct, clocks transported from A to B won't, in general, be synchronized at B even if they were at A. For when in motion from A to B, they will, in general, be traveling at different speeds with respect to the aether and, hence, suffering different amounts of "slowing down." Clearly, the right clock to use to determine synchronization of clocks at A and B will be one moved very slowly with respect to the aether and hence suffering minimal distortion as it is moved. But to know which clock that is, we would have to know which state of motion was the rest state in the aether, which is what we were trying to determine in the first place!

Suppose we knew which state of motion was at rest in the aether. Because light, relative to the aether, travels with the same speed in all directions, an easy way to synchronize clocks at A and B would be to send a light signal from A that was reflected at B and returned to A. As the light took the same time to get from A to B and from B to A, the event at A simultaneous with the reflection at B could be taken to be the event at A midway in time between the emission and reception of the light signal at A as determined by a clock at rest at A. But, says Einstein, as far as the round-trip experiments go, it is as if light had this same speed in all directions no matter what is the uniform state of motion of the observer. Assume light really does travel at the same speed relative to any observer in uniform motion. Then each such observer can use the reflected-light method to determine which events take place at the same time as which other events.

It is easy to see that taking this as our definition of simultaneity for distant events will result in observers' disagreeing about *which* pairs of events take place at the same time, as can be seen from Figure 2.2 and its explanation. Well, which observer is *right* in his attributions of simultaneity? According to the aether theory, only the observer at rest in the aether. The others are being deceived by their taking light to travel at the same speed in all directions relative to their laboratories, when it really does not. According to Einstein, all the observers are correct in their attributions of simultaneity. It is just that there is no such thing as "occurring at the same time," only "occurring at the same time *relative* to a particular state of uniform motion." We can reconcile the null results of the round-trip experiments with the Galilean assumption that all uniformly moving observers see the same physical

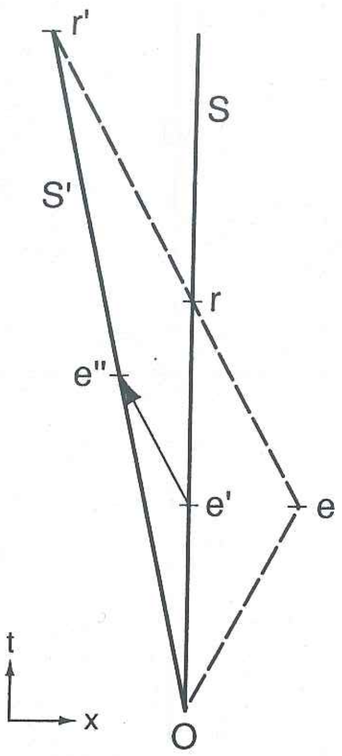

Figure 2.2 The Einstein simultaneity definition and the relativity of simultaneity. *OS* represents the events in the life history of one observer, an observer who stays at a constant *x* position as time, *t*, goes on. *OS'* the life history of another observer moving (relative to *OS*) to the left. Because *e'* is halfway in time from *O* to *r*, the events of emitting and receiving a light beam reflected at event *e*, *S*, taking the speed of light to be the same toward and back from *e*, takes *e'* to be simultaneous with *e*. *S'*, reasoning similarly, takes *e"* to be simultaneous with *e* because it is halfway in time from *O* to *r'*. Yet because a causal signal can leave *e'* and arrive at *e"*, both *S* and *S'* agree that *e'* and *e"* cannot be simultaneous. In relativity, events are or are not simultaneous only relative to a chosen "inertial frame of motion" like that of *S* or that of *S'*.

phenomena by simply dropping the intuitive notion that there is an absolute, nonrelative notion of "occurring at the same time."

We can mitigate some of the strangeness of this conclusion if we look at the concept of "being in the same place." Imagine two observers in motion with respect to one another. The first observer is hit on the head at two different times. Did the blows occur "at the same place"? "Yes," says the struck observer, "they both occurred in the place where the top of my head was located." "No," says the other observer, "one occurred near to me and the other far away." Which claim is correct? Unless one believes in Newton's "space itself," relative to which one

and only one of the observers can really be at rest, why not say that "at the same place" is just a relative notion? Two events can be at the same place relative to one observer and at different places relative to another in motion with respect to the first. And, if Einstein is correct, it is just the same with "at the same time."

To get the full picture of space and time proposed by Einstein requires one further assumption. It involves the claim that all places and directions in space and time are alike but goes beyond this in making an assumption that amounts to a posit that the spatiotemporal structure of the world is "flat." We will examine this notion of "flatness" in more detail in "Gravity and the Curvature of Spacetime." The assumption needed is the linearity of the relations of spatial and temporal separations for one observer with respect to those of another observer. With this additional posit, a structure of space and time is constructed in which observers in motion with respect to each other will attribute quite different spatial separations of events from each other and will also attribute quite different temporal separations between the events. The spatial and temporal separations attributed to a pair of events by one observer can, however, be calculated from those attributed to the pair by another observer moving with respect to the first, by means of the so-called Lorentz transformations, formulas originally derived in the context of the earlier compensatory theories.

Although the spatial and temporal distances between two events will vary from observer to observer, it is important to note that a consequence of the basic postulates of the theory is that another quantity, the so-called square of the interval between the events, will have an invariant value: It will be the same for all uniformly moving observers. It can be calculated from the time separation between the events in one observer's frame, t, and the spatial separation in that same reference frame, x, and the velocity of light, c, by means of the formula: $I^2 = x^2 - c^2t^2$. Whereas t and x will vary from observer to observer, I^2 will remain the same for all of them. A crucial step in this proof relies on the fact that all observers are attributing to light the same invariant speed, c.

Minkowski Spacetime

All the consequences of Einstein's theory for a new conceptualization of space and time can be summarized in the notion of Minkowski spacetime, the arena of all physical processes in the theory of special relativity. The basic idea here is to start with point-event locations, as the fundamental constituents out of which spacetime is built. One can think of these as the possible locations of happenings that are instantaneous and unextended spatially. These point-events take the place of the spatial points and moments of time of the prerelativistic theory. It is the basic structures imposed on the set of these spacetime points, the events or event locations, that constitute the framework of the new picture of space and time. (See Figure 2.3.)

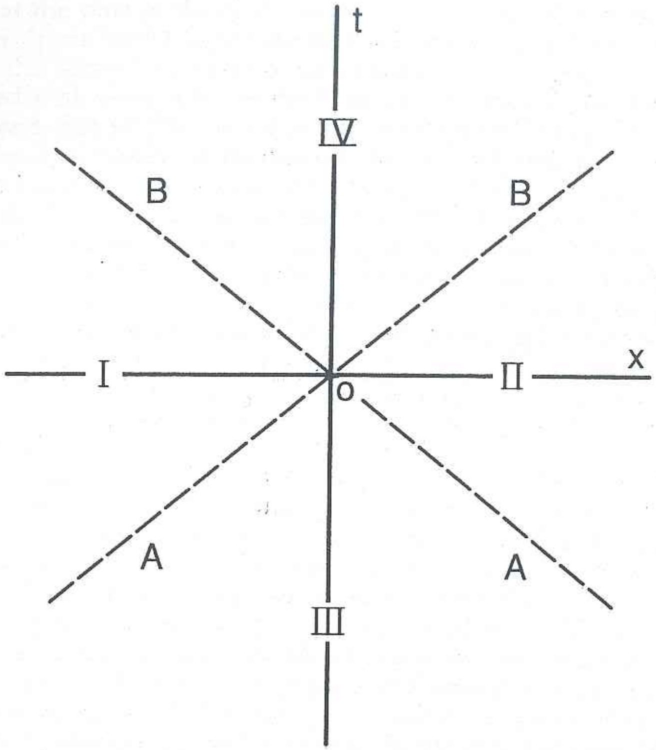

Figure 2.3 Some elements of Minkowski spacetime. The line *t* represents an inertial observer with *o* an event in that observer's life. The line *x* represents the events simultaneous with *o* for the observer. *A* and *B* represent light signals coming from the past to *o* and leaving *o* into the future. The events in regions *I* and *II* are so far from *o* in space and close to it in time that a signal would have to travel faster than light to connect such an event to event *o*. It is generally assumed that there are no such signals. The events in regions *III* and *IV* are events connectable to event *o* by causal signals traveling at less than the speed of light.

Pairs of these event locations have a definite interval between them, invariant and absolute in the structure. For a given observer in a particular state of uniform motion, a definite spatial separation and a definite time interval between the events can be derived, but the values of these are relative to the particular observer's state of motion.

Two events whose interval separation has value zero are such that a light signal in vacuum emitted at one place-time could arrive at the other place-time. Notice that "interval" is unlike spatial distance in that distinct events can have zero-interval separation. Such events are said to have null or lightlike separation. Events whose interval squared is negative are close enough together in space and far enough apart in time so that signals propagating slower than light can get from one to the other. They are said to have timelike separation. Event pairs whose interval squared is positive are too far apart in space and too close together in time for any signal traveling at the speed of light or less to connect them. If we assume that light is the limiting fastest signal, the events are unconnectable by any causal process whatever and are said to have spacelike separation. If we pick one event as origin, the class of events null separated from it divides the spacetime into interior and exterior regions of events timelike and events spacelike separated from the origin event. This separating class of events lightlike separated from the origin event consists of a future and past component. Together these are called the "light cones" of the origin event. (Actually they are cones only in a spacetime of two rather than the actual three spatial dimensions.)

In the usual flat space of Euclidean geometry, straight lines exist. Minkowski spacetime also has straight-line paths. If the intervals between the points on the geodesic path are spacelike, the path represents a straight spatial line. The latter is a straight line in the space at a time generated from the spacetime by picking some uniformly moving observer and taking as space at a time for him a collection of events all of which are simultaneous in his reference frame. Straight lines whose events have null separation represent the paths of light rays traveling in a vacuum. Timelike straight lines represent the path through space in time of some particle in uniform motion.

On a diagram we could represent some observer at rest in a frame of uniform motion by a vertical straight line. Any other uniformly moving observer who coincides with our first observer at the origin event would be represented by a straight line at an angle to the vertical. It is important to recognize that which line is vertical carries no physical significance. Only if we had a Newtonian notion of who is really at rest in space itself would there be some real significance to representing one observer as always at the same place and the other uniformly moving observers as changing places over time. But Minkowski spacetime has no such notion of which uniformly moving observer has zero real velocity, for all uniform velocities are physically on a par in this spacetime picture.

Having chosen some uniformly moving observer, we can also represent on the diagram by a straight line all of those events simultaneous to the origin event relative to that observer's state of motion. Diagrammatically, this straight line really represents "space at a time" for the observer, which is, of course, three dimensional. But we must suppress two spatial dimensions to get the diagram on a plane; therefore, a whole infinite, flat, Euclidean three-dimensional "space at a time" is represented by a line. For the observer in motion with respect to our first observer, a different straight line will represent all the events simultaneous to the origin event relative to this new observer's state of motion. A different line is needed because for the two observers different events are classed as simultaneous to the origin event, and what counts as space at the time of the origin event depends on an observer's state of motion. It can easily be shown that the simultaneity line (space at a time) for the second observer represented in the diagram would have to be tilted with respect to the first observer's simultaneity line.

We noted that in the compensatory theories originally designed to explain the null results of the round-trip experiments, it was posited that objects in motion with respect to the aether shrank in their lengths, and that clocks in motion with respect to the aether slowed down. In Minkowski spacetime, there is, of course, no aether. Yet length contraction and time dilation do occur. Let a meter-long stick be at rest in a uniform motion reference frame. That meter stick will be declared to have a length less than one meter in any other uniformly moving frame. Let a clock be at rest in one uniformly moving frame. That clock will be declared to be "running slow," i.e., taking more than one second to tick off one second on its face, by an observer in any other uniformly moving frame.

What is striking is that this contraction of length and dilation of time is perfectly symmetrical. Meter sticks at rest in your reference frame are taken as shortened by me (the two of us in relative motion), but meter sticks in my frame are taken as shortened by you. And the slowing down of clocks is equally symmetrical. Despite the appearance of inconsistency here, there is none, for length and time interval are now relative to an observer, and the assertions made are perfectly consistent. Direct evidence of the real existence of these phenomena is available, for example, in the life span of unstable particles—inexplicably overlong in prerelativistic terms—created in the upper atmosphere and observed on the surface of the earth. Only the relative slowing down of their decay process, because of their high velocity relative to us, can account for the phenomenon.

This consequence of relativity gives rise to a wide variety of paradoxes, apparent contradictions that really aren't contradictions, some of which can be found in any standard text on relativity. For example, a man carrying a pole runs into one end of a barn and out the other. When the pole is at rest with respect to the barn, it has the same length as the barn. Because the pole in motion is shorter than the barn, someone can close both doors on the runner while he and the pole are in the barn. But to the runner the barn is shorter than the pole, so this is clearly impossible. The key is to think about the time order in which processes occur from the differing perspectives of the runner and of the observer at rest on the barn. For the man at rest on the barn, both doors are shut while the runner is in the barn with the pole. The runner sees the farther door open and his pole protrude from the barn before the nearer door is ever closed behind him.

The spacetime of special relativity, Minkowski spacetime, requires us to make another distinction about time that doesn't occur in the prerelativistic theory. We have noted that any observer will attribute a certain time interval between two events, and that this interval will

vary from observer to observer. This is called the coordinate time interval between the events relative to the observer in question. Another notion of time arises when we consider someone who moves from one event (one place at a time) to some other event (a different place at a different time) along some spacetime path, through a succession of places-at-a-time. Let this agent carry a clock with him set at zero at the first event. This clock will read a definite value at the final event. Surely all observers will agree on what that value is because the coincidence of the clock's reading that value with the final event will be agreed upon by all, as these are events at the same place and there is no relativity of simultaneity in that case. This time is called the proper time between the two events.

But the elapsed proper time between two events will vary depending upon the spacetime path by which the clock is carried from one event to the other. This phenomenon is without precedent in prerelativistic physics. In fact, it can easily be shown that the time elapsed on a clock carried from one event to another will be maximal if the path followed from the first event to the second is one of unaccelerated, uniform motion. This is the source of the famous twin paradox, according to which if one twin remains in a frame of uniform motion while the other takes a course through space and time that involves accelerated motion but that brings him back into coincidence with his stay-at-home twin, the adventurous twin will be younger—will have shown, for example, less biological aging—than his twin when the two meet once again. Evidence that this consequence of relativity is real comes from unstable particles sent around the circular paths of accelerators. Fewer of them decay than their compatriot particles in a group remaining at rest in the laboratory between the first moment when they coincide and the second moment at which they coincide. As usual there is no contradiction in the theory here, just phenomena we hadn't expected, a result of the surprising nature of spacetime. (See Figure 2.4.)

We noted that Newtonian mechanics conformed to Galileo's principle that all physical phenomena would appear the same to any observer in a state of uniform motion, although the fact that one's laboratory was in accelerated motion would reveal itself in observable consequences. The old theory of mechanics, when put in the new, relativistic spacetime, no longer would satisfy that principle. Hence, a new mechanics was developed by Einstein that reconciles the Galilean relativity of mechanical phenomena with the new spacetime picture. The source of this theory is simple. The older mechanics obeyed such principles as the conservation of energy, the conservation of momentum, and the conservation of angular momentum. These are now realized to be the consequences of fundamental symmetries of the spacetime structure (in particular, the fact that all spacetime points are structurally alike, as are all spacetime directions). These symmetries obtain in the new spacetime as well, so we may hold to the old conservation rules and derive the new mechanics from them. In the new me-

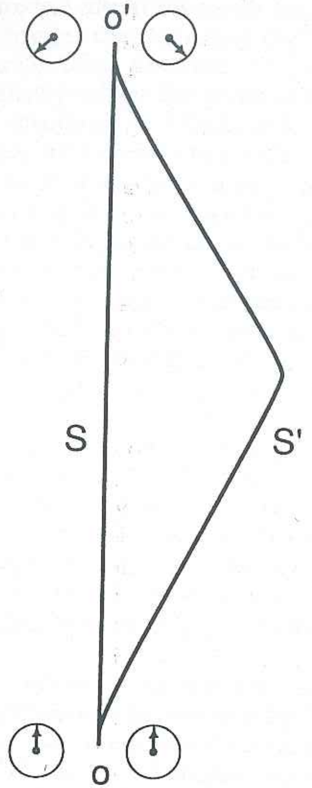

Figure 2.4 The twin paradox. *S* is an observer who remains inertial and carries a clock from event *o* to event *o'*. The time elapsed on the clock is represented by the left-hand clock faces. *S'*, originally at rest with respect to *S*, accelerates off to the right, travels to the right at a uniform speed, reverses direction of relative motion, returns to *S*'s location and again accelerates so as to come to rest relative to *S* at the location *o'*. The elapsed time on a clock carried by *S'* is represented by the right-hand clock faces. Special relativity predicts that less time will have elapsed on the clock carried by *S'* along the accelerated path from *o* to *o'* than will have elapsed on *S*'s clock.

chanics is found the famous consequence of relativity, for example, of the equivalence of mass and energy, i.e., that the more kinetic energy an object possesses, the greater will be its resistance to further acceleration by a force.

We also noted that Einstein assumed that the speed of light in a vacuum was a maximal speed of propagation of any signal whatever. Such a posit fits in nicely with the new spacetime picture.

We can, for example, find pairs of events, *A* and *B*, and observers *O1* and *O2*, such that *A* is before *B* relative to *O1* and *B* is before *A* relative to *O2*. But these will always be events that have spacelike sep-

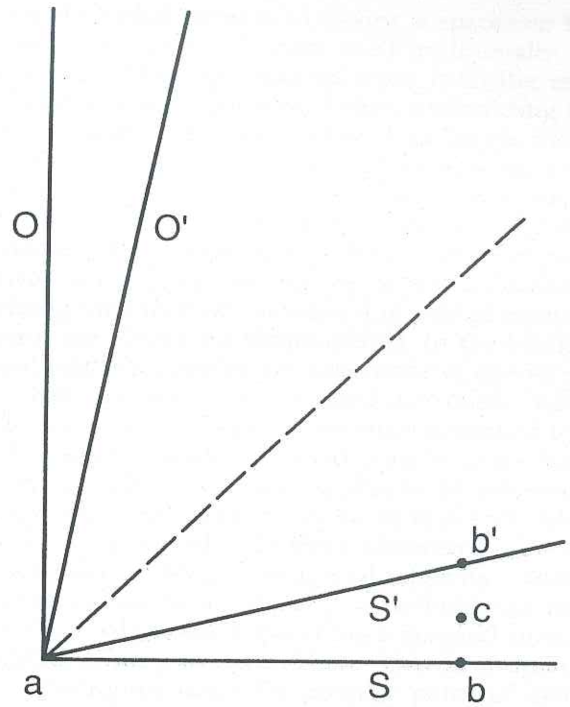

Figure 2.5 The relativity of time order of events in special relativity. *O* is an inertial observer. *O'* is another inertial observer moving to the right relative to *O*. The line *S* is the class of events *O* takes as simultaneous with event *a*. The line *S'* is the class of events *O'* takes as simultaneous with *a*. For *O* event *c* is after event *b* and hence after event *a*. But for *O'*, event *c* is before event *b'* and hence before event *a*. Such a reversal of time order can occur only for events, like event *a* and event *c*, that are not causally connectable to one another.

aration. This means, assuming the limiting velocity of light, that events will have different order in time relative to two observers only if the events are not connectable by any causal signal whatever. Events that are connectable by a causal signal, traveling at or at less than the speed of light, will appear in the same time order to all observers, although the amount of time between them will vary from observer to observer. (See Figure 2.5.)

It has been pointed out that one need not hold to the posit of the limiting velocity of light in order to have a consistent theory in which spacetime is Minkowski spacetime, the spacetime of special relativity. If one simply insists that conditions of the world are such that causal paradoxes are avoided, one can tolerate "tachyons," causal signals at higher than light speeds. The consistency constraint is needed because positing tachyons in Minkowski spacetime would allow for closed causal loops, in which one event causes itself. If initial conditions could be freely chosen, a paradoxical situation could be generated (I shoot myself before pulling the trigger that launches the bullet). No such above-luminal speeds have ever been detected, however, and the standard versions of special relativity adopt the posit of light as maximal causal signal along with the structure of Minkowski spacetime with its invariant speed of light for all inertial observers.

Nothing in the spacetime of special relativity, as we have noted, plays the full role of Newton's space. For Newton, space itself provided a genuine standard of what it is for an object to be really at rest, even if no empirical consequences arose from uniform motion with respect to space itself. In Minkowski spacetime, nothing provides a standard of when two events that are not simultaneous with each other are "at the same place." It is therefore meaningless to ask whether an object remains at one and the same place through time, although it is perfectly meaningful to ask whether an object's *relative* position, that is, position with respect to some other material objects taken as a frame of reference, remains unchanged over time. But the distinction between being genuinely in uniform motion or not does remain in this spacetime. Whether the path of some material particle through the spacetime, the timelike path that represents the succession of place-times the object occupies, is a straight line or not, that is, whether it is one of the timelike paths that is a geodesic of the spacetime, is a perfectly meaningful question.

The distinction, then, between an object's being in uniform motion or being in accelerated motion—represented by a curved timelike path in the spacetime diagram—remains absolute in the sense that this distinction has nothing to do with the motion of the object in question relative to other material objects. Instead, this distinction is determined by an object's motion relative to the structures of the spacetime itself. In Newtonian physics, genuinely accelerated motion revealed itself by the presence (in the accelerated laboratory) of inertial forces, acting on objects and generated, allegedly, by the acceleration of the objects relative to space itself. In special relativity, real acceleration shows up in this way and in other ways as well. We noted, for example, that it was only when one of the round-trip experiments was performed with light in a laboratory in uniform motion that the null results were obtained. In an accelerated apparatus the light will take times around the paths that reveal the existence of the absolute acceleration of the experimental device. Although there is no such thing, then, as "being in the same place" in any sense other than relative to some material standard, there is, in special relativity, as much real significance to "being in uniform motion" in an absolute sense as there is in the Newtonian theory.

Neo-Newtonian Spacetime

Once the Minkowski spacetime of special relativity had been constructed, it was noted that one could go back and construct a spacetime

appropriate for the earlier Newtonian theory, a spacetime that had some advantages over the notion of space itself traditionally postulated in Newtonian physics. The main insights come from the realization that taking event locations as primitive and then constructing the spacetime by imposing structure on the set of event locations is the best systematic route for constructing a spacetime appropriate for what are taken to be the observable quantities posited by any given theory.

In Newtonian physics, the notion of simultaneity for distant events is presupposed as an absolute notion. So to construct our new spacetime for Newtonian physics, we impose on the collection of event locations a definite time interval between any pair of events. When this interval is zero, the events are simultaneous. In the Minkowski spacetime of special relativity, spaces are collections of events simultaneous relative to a given observer. It is assumed that these "relative" spaces have the ordinary three-dimensional structure described by Euclid's geometry. In the revised Newtonian spacetime, with its absolute notion of simultaneity, we can, again, take spaces to be collections of simultaneous events. Thus each event will be in one and only one space, and space is, again, assumed to be three-dimensional Euclidean space.

In the Newtonian context, as in special relativity, what counts as a path of uniform motion of an object is a well-defined notion. So we impose on this new Newtonian spacetime a demand similar to the one imposed on Minkowski spacetime: There must be a definite notion of straight-line paths representing the possible paths of motion through space in time of freely and uniformly moving particles. Now Newton assumed that there was such a thing as one event's being "in the same place" as some other nonsimultaneous event. If we impose that structure, a definite notion of same place for nonsimultaneous events, on the spacetime we are constructing, we build up Newton's absolute space picture of space and time. But this would give us features of the world without empirical consequences, such as the magnitude of the uniform speed of an object with respect to "space itself." If we leave that "same place at different times" structure out, however, we obtain a new spacetime, sometimes called Galilean spacetime, sometimes called neo-Newtonian spacetime. In this spacetime, absolute uniform motion is well defined, but absolute sameness of place through time is not.

In this new spacetime picture, absolute accelerations exist and have observable consequences, but there is no such thing as the absolute velocity of an object. This is just what we want. The physicists' need for a new approach to space and time in order to confront the startling and puzzling results of the optical round-trip experiments led to deep insights into what the components were of the picture of space and time that we held intuitively and that, in a refined version, underpinned the physical picture of the world of Newtonian science. By confronting the new experimental facts and constructing the conceptual apparatus to do justice to them, physicists came up with new ways of looking at possible theories to account for the older posited observational facts. As we shall see, the existence of these new structures for describing and explaining the spatiotemporal features of the world had an important effect on our philosophical understanding of the nature of space and time and of our access to knowledge about their nature as well. But before taking up those issues, we will look at a second revolutionary change in our views about the nature of space and time, once again generated out of the fertile scientific imagination of Einstein.

Gravity and the Curvature of Spacetime

Gravity and Relativity

In his greatest work, the *Principia*, Newton proposed a theory that would, among other things, explain the motion of the planets around the sun in the elliptical orbits that had been so carefully described by J. Kepler. The theory accounting for this motion has two components. One is Newton's theory of dynamics, his general theory relating motions to the forces acting upon the objects in motion. Based on a background assumption of absolute space and a definite absolute rate of time, the theory incorporates Galileo's principle that objects not acted upon by any forces remain in a constant state of uniform motion. It then posits that change of motion (acceleration) will be proportional to the forces acting upon a body and inversely proportional to the intrinsic propensity of a body to resist changes of motion, its so-called inertial mass.

The other component of Newton's theory concerns the force responsible for the observed motions of astronomical bodies (and for many other phenomena, such as the way in which bodies fall toward the surface of the earth and the tides). Once again building upon Galileo's important observation that, air resistance to the side, all objects suffer a uniform acceleration toward the earth when they are in free-fall near the earth's surface, Newton posits a general force of gravity acting between all material objects. Gravity is always an attractive force. The magnitude of the force exerted between the bodies is taken to be proportional to the inertial mass of each body and inversely proportional to the square of the distance between them. Newton's Third Law of Motion asserts that the force exerted by the first body on the second will be matched by a force of equal strength—but oppositely directed—exerted by the second body on the first.

The facts that the force increases in proportion to the inertial mass, but that the resistance of the body to acceleration is also proportional to the inertial mass, immediately produce Galileo's result that all bodies accelerate equally when subjected to the gravitational force exerted by some fixed body if the test objects are at the same location relative to the object exerting the gravitational force. Newton demonstrated that the combination of the laws of dynamics and the law of gravitational force that he postulates will lead to Kepler's laws of planetary motion or, rather, to a slightly corrected version of them.

It should be no surprise, then, that Einstein, having demonstrated the necessity for a new dynamical system and having constructed one consistent with the new spacetime of special relativity, takes up the problem of the construction of a new theory of gravity. This theory, which is clearly needed, must be consistent with the new spacetime ideas. Newton's theory, for example, takes the gravitational interaction between bodies to be instantaneous, but relativity takes all signals to propagate at a speed less than or equal to that of light. Quite a variety of alternatives to the Newtonian theory can be constructed that are compatible with the new relativistic spacetime. Indeed, a continuing program in experimental physics amounts to testing these alternatives against one another, looking for possible observations to rule out some of the possibilities. But the novel gravitational theory that has stood up best against experiment, and the one of greatest theoretical elegance, is Einstein's own. This is called the general theory of relativity. It is also the theory that posits to the world a nature that is of great interest to philosophers. I will spend the rest of this section sketching some of the ideas that led Einstein to this novel theory of gravity, which, as we shall see, constitutes a novel theory of the structure of spacetime itself. I shall outline some of the basic components of the theory and explore a few of its consequences that are of importance for the philosopher.

Einstein begins with Galileo's observation that the acceleration induced in an object by gravity is independent of the object's size and of what it is made of. Gravity is unlike any other force in having this universal effect. Consider the case in which the object is forced into acceleration by a distant enough gravitating object so that the gravitational field is effectively constant within the laboratory. Einstein observes that a small test object in a laboratory would accelerate relative to that laboratory in just the same manner as it would if no force were acting on the test object but, instead, the laboratory itself were being uniformly accelerated in the opposite direction to that of the particle's acceleration. In the latter case, any test object of any mass or composition would appear to accelerate uniformly with respect to the laboratory. It is the universality of gravity that allows us to replace the gravitational force by a reference frame acceleration.

Perhaps, Einstein suggests, all the effects of gravity could be duplicated by such a laboratory acceleration. This leads to the hypothesis that gravity will have effects on things other than particulate matter. If we shine a light beam across a laboratory in accelerated motion we expect the beam to follow a path that is not a straight line relative to the laboratory. Shouldn't gravity, then, deflect beams of light that pass near a gravitating body?

Perhaps more surprising is the conclusion that we ought to expect gravity to have an effect on measurements of time and space intervals, as revealed by idealized clocks and measuring rods. The argument in favor of the temporal effect is the easier one to construct and follow. Imagine an accelerated laboratory with a clock at its upper end and an identical clock at its lower end. Signals are sent from the lower clock to the upper and the rate of emission of the signals, as determined by the lower clock, and of their reception, as determined by the upper, are compared. By the time a signal released from the bottom has gotten to the top, the top clock is in motion with respect to the uniformly moving reference frame in which the bottom clock was at rest when the signal was released. Arguing either from the time-dilation effect of special relativity or from the so-called Doppler effect, which, even pre-relativistically, shows that a signal released from a source with a given frequency will appear to have a lower frequency when observed by someone relative to whom the source is in motion, it becomes plausible to claim that the lower clock will appear to be running slow as determined by the upper clock. That is, the frequency with which the signal is received by the upper clock is lower than that with which it is emitted as determined by the lower one. (See Figure 2.6.)

But now consider the laboratory not accelerated, with the apparatus all located at rest in a gravitational field. By Einstein's argument (often called the Principle of Equivalence), we ought to expect that the clock lower down in the gravitational field will appear, to the clock located higher up, to be running slow. Notice that this has nothing to do with the gravitational *force* felt by the two clocks but, rather, is determined by how much further down the gravitational "hill" one clock is than the other. So we ought to expect gravity to have an effect on our measurement of time intervals. Similar, but somewhat more complicated, arguments can be given that lead us to expect gravity to affect spatial measurements as well.

Taken together, these arguments led Einstein to the astonishing suggestion that the way to deal with gravity in a relativistic context was to treat it not as some force field acting in spacetime but, instead, as a modification of the very geometric structure of spacetime. In the presence of gravity, he argued, spacetime is not "flat" but is "curved." To know what that means, however, we must look briefly at the history of geometry as treated by the mathematicians.

Non-Euclidean Geometry

Standard geometry as formalized by Euclid derived all the geometric truths from a small set of allegedly self-evident, basic postulates. Although Euclid's axiomatization of geometry is not, actually, complete (i.e., sufficient in itself to allow all the derivations to be carried out without presupposing other underlying and hidden premises), it can be so completed. For a long period of time, puzzlement existed about Euclid's so-called Parallel Postulate. It is equivalent to the claim that through a point not on a line, one and only one line can be drawn that is in the common plane of given line and point and that will not intersect the given line in either direction no matter how far the lines are extended. It seemed to the geometers that this postulate lacked the self-

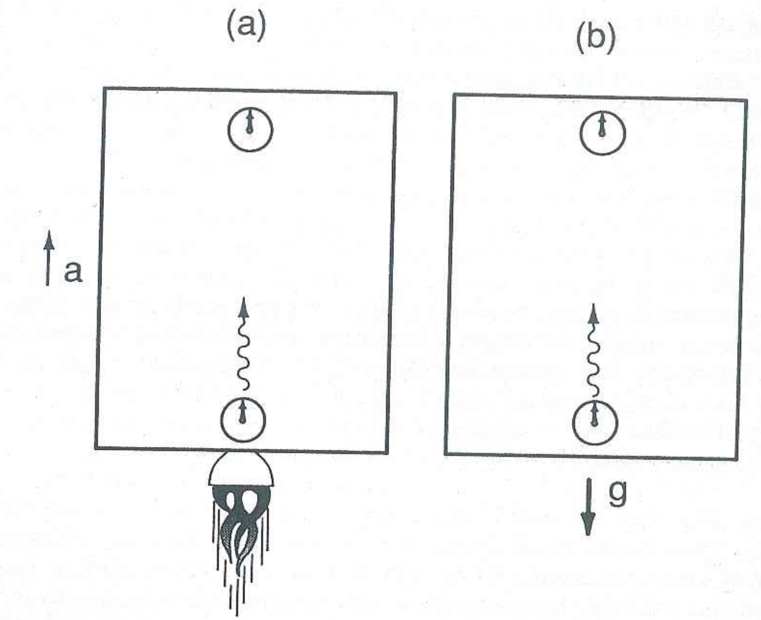

Figure 2.6 The gravitational red shift. (a) represents an accelerated laboratory with a clock on the floor and one attached to the ceiling. Because a signal emitted from the floor clock is received at the ceiling clock when the laboratory is moving with a velocity relative to the frame of motion in which the signal was emitted (because of the acceleration of the laboratory), the floor clock will be recorded as "running slow" by the ceiling clock in much the manner a whistle moving away from an observer is heard by that observer with a lower pitch than it would be if the whistle were stationary with respect to the observer. General relativity posits that a similar result will be obtained in a laboratory not accelerated but fixed in a gravitational field—as in (b). A clock lower down in the gravitational field will be recorded as "running slow" by a clock higher up in the gravitational potential. This is called the gravitational red shift. It indicates one way in which gravity can be taken to affect the metric structure of spacetime.

evidence of the other, simpler hypotheses (such as "Equals added to equals are equal," and "Two points determine a straight line between them"). Could this "suspect" postulate be derived from the other postulates, making it unnecessary as an independent assumption? If one could show that the denial of the Parallel Postulate was inconsistent with the other postulates, one could show this derivation to hold by the method of *reductio ad absurdum*. But could this be shown?

Denying the Parallel Postulate can go in two directions. The postulate says that one and only one parallel line through the point exists, and to deny that, one could affirm that no such parallel line existed or that more than one did. In 1733 G. Saccheri showed that the No Parallels Postulate was, in fact, inconsistent with the remaining axioms, at least if these were understood in their usual way. He was unable to show that the Many Parallels denial was so inconsistent. By the nineteenth century, J. Bolyai, N. I. Lobachevsky, and K. F. Gauss had realized that one could construct consistent geometries that adopted Euclid's postulates but that had a Many Parallels Postulate in place of Euclid's Parallel Postulate. B. Riemann then showed that if the other axioms were slightly reinterpreted, a new geometry with a No Parallels Postulate replacing Euclid's Parallel Postulate could be constructed that was also logically consistent. The reinterpretations needed are that "Two points determine a straight line," must be read so that sometimes more than one straight line contained a given pair of points and "A line may be extended arbitrarily in both directions," must be read to assert that a line wouldn't meet an end point if extended but not to imply that a fully extended line had infinite length.

Later it was realized that when these new non-Euclidean geometries were taken to be two-dimensional, plane, geometries, they could be understood in a Euclidean fashion as the geometry of shortest distance curves (geodesics) on curved two-dimensional surfaces. In particular, Riemannian axiomatic geometry was just the geometry of figures constructed by arcs of great circles on the surface of a sphere. But what could the logically consistent three-dimensional non-Euclidean geometries be taken to be about, or were they, even if logically consistent, absurd for other reasons?

Gauss carried geometry further by developing a general theory of arbitrarily curved two-dimensional surfaces. These are characterized by a number—known as the Gaussian curvature—at each point. How this curvature varies with distance as measured along curves drawn in the surface determines the shape of the curved surface. Gauss thought of these curved surfaces as embedded in ordinary Euclidean three-dimensional space. An important result of his work, however, was that one could characterize some of the aspects of curvature ("intrinsic" curvature) by means of quantities that could be determined by an imagined two-dimensional creature confined to the curved surface and not even aware that the embedding three-dimensional space existed. From this new perspective, it turned out that the geometries described by the older axiom systems could be understood as special cases. Euclidean two-dimensional geometry, the geometry of the plane, is the geometry of the surface whose Gaussian curvature is everywhere zero. Riemannian geometry, the geometry of the two-dimensional surfaces of spheres, is just the geometry of a surface whose Gaussian curvature is constant and positive. Lobachevsky-Bolyai geometry is the geometry of a two-dimensional surface whose Gaussian curvature is the same at each point and negative. Negative curvature characterizes a point like that in the center of a mountain pass at which the surface curves "in opposite directions" along different paths through it.

Riemann then went on and generalized Gauss's theory of curved surfaces to spaces of any dimension whatever. Whereas Gauss presupposed that the surfaces in question were embedded in a flat Euclidean space, Riemann made no such assumption. After all, it was a result of Gauss's work that some aspects of curvature would be available to a two-dimensional creature ignorant of the embedding space. General Riemannian geometry deals with these aspects of curvature, the intrinsic aspects. (This general Riemannian geometry of curved n-dimensional spaces is not to be confused with the earlier axiomatic Riemannian geometry.) The basic assumption of this geometry is that the curved n-dimensional space is approximable in small enough regions by a Euclidean, flat, n-dimensional space. For curved surfaces in flat three-dimensional space, these approximating surfaces can be represented as planes tangent to the curved surface at a point; the planes are also located in the embedding three-dimensional space. For a general Riemannian curved n-dimensional space, these "tangent planes" are posited to exist only in the sense that as far as intrinsic n-dimensional features go, the n-dimensional curved space can be approximated at a point by a flat, n-dimensional Euclidean space.

What are some aspects of curved spaces? How, for example, could a three-dimensional creature living in a curved three-dimensional space find out that the space was, in fact, curved? Intrinsic curvature reveals itself in distance measurements. An n-dimensional creature can make enough distance measurements between points to assure itself that there was no way these points could be located in a flat n-dimensional space and have the minimal distances between them along curves that the creature's points do. For example, a check of shortest airline distances between cities on the Earth could tell us that the Earth had not a plane surface but, instead, a surface approximating that of a sphere. In a curved n-dimensional space, shortest-distance curves, called the geodesics of the space, fail to be the straight lines that they would be were the space flat. These lines are also the lines of "least curvature" in the space. Intuitively, this means that the lines, although they cannot be straight, given the curvature of the space, deviate from straightness no more than they are forced to by the curvature of the space itself.

Curvature can reveal itself in other ways as well. For example, if we take a directed line (a vector) and move it around a closed curve in a flat space, all the while keeping it as parallel to itself as possible as we move it, when we return to the point of origin, the vector will point in the same direction at that point as when we began. But in a curved space, such parallel transport of a vector around a closed loop will, in general, change the direction of the vector so that it will point in a direction at the end of the transport that is different from its original direction when the journey started.

A flat three-dimensional space is of infinite extent and has infinite volume. A Euclidean plane is of infinite extent and has infinite area. But the intrinsically curved surface of a sphere, although it has no boundaries, has a finite area. A two-dimensional creature living on a spherical surface could paint the surface. It would never encounter a boundary to the surface. But after a finite time the job would be done, with the whole surface painted. Similarly, a three-dimensional creature living in the three-dimensional curved space that is analogous to the spherical surface, living in a so-called three-sphere, could fill the region with foamed plastic. Although never encountering a boundary wall to the space, it would, in a finite time, finish the job, with all the volume of the three-dimensional space filled by a finite amount of plastic foam.

It seems clear, then, that the notion of a curved *n*-dimensional space, including a curved three-dimensional space, is not only logically consistent but manifestly nonabsurd. As long as we are sticking to intrinsic features of curvature, we are not making the assumption that the space is sitting in some higher-dimensional, flat embedding space. And the features of curvature intrinsic to the space are manifestly ascertainable by straightforward techniques to a creature living in the space. Could it not be the case, then, that the actual three-dimensional space of our world was curved and not the flat space characterized by the basic postulates of three-dimensional Euclidean geometry? Such speculations naturally accompanied the discovery of the new geometries.

Using Non-Euclidean Geometries in Physics

There was some speculation in the nineteenth century about the possible reality of curved space. W. Clifford, for example, suggested that it was conceivable that matter was actually little regions of highly curved space in a three-dimensional space that was flat in the large. It was clear that a large-scale curvature of space could be detected only on the largest, astronomical, scales, for generations of experience had shown us how well Euclidean flat three-dimensional geometry worked in our descriptions of the world. Certainly it worked well for measurements of the ordinary sort and even in the description of such things as the structure of the solar system.

It was only with Einstein's new relativistic theory of gravity, however, the general theory of relativity, that curved geometry became an essential part of a plausible physical theory. We have seen that one could argue with plausibility that gravity would affect all objects dynamically in the same way, independently of their size and constitution. Thus a material object, which would, in the absence of gravity or other forces, follow a path of uniform direction and speed would, in the presence of gravity, follow a different path. But the change in this path would depend only upon the gravitational field and the initial place and velocity of the object. It wouldn't depend on the mass of the object or the material out of which the object was made. It is this independence of the effect of gravity on the object's size and structure that makes a "geometrization" of the gravitational field possible.

When combined with the arguments in favor of a gravitational effect on metric features of the world as determined by rods and clocks, the idea of treating gravity as curvature becomes plausible. But it is not, at

least fundamentally, curved space that Einstein posits but, instead, curved spacetime. In the Minkowski spacetime of special relativity, free particles traveled timelike straight lines, the timelike geodesics of the spacetime. Now, Einstein suggests, we are to think of particles that are acted upon only by gravity as "free" particles, traveling, not timelike straight lines, but the curved timelike geodesics of a spacetime that is curved. A fundamental result of Riemann's geometry is that through a point in a given direction there passes a single geodesic path. In Riemannian spaces, the geodesics are both the paths of minimum curvature and of (locally) shortest distance. With the new metric of spacetime, it is best to think of the "least curvature" definition of geodesics as the fundamental one. In spacetime, if one specifies a direction at a point one will simultaneously specify a spatial direction and a speed in each direction. So the timelike geodesic through a point in a given direction will correspond to specifying the initial place and the initial velocity of a particle. The path specified by the geodesic will then be unique. And this is just what we want for gravity because, given an initial place and velocity, the path of any particle in a gravitational field is the same. Light, which in special relativity follows the straight-line null geodesics of Minkowski spacetime, is now taken to follow null geodesics in the curved spacetime, geodesics that will, in general, not be straight lines. (See Figure 2.7.)

One could determine the curvature of a spacetime by following out the paths of the "free" particles and light rays, that is, the particles and light rays acting only under the influence of gravitation, now taken to be simply the curvature of the spacetime. But one could also, at least in principle, determine the curvature structure by making enough measurements of spatial and temporal separations of events and combining these measurements into the interval separation, which is the metric for spacetime. General relativity posits that the spacetime so determined will agree with that determined by following out the geodesic motions of particles and light rays, the clocks and rods being used to make the temporal and spatial measurements also being affected by the gravitational field, in the sense that they properly measure these metric qualities in the curved spacetime.

Traditional gravitational theory had two parts: One specified the action of gravity on test objects; the other specified the kind of gravitational field that would be generated by a source of gravity. In the older theory, gravity was a force that accelerated all material objects at a place in the gravitational field to the same degree. In the newer theory, gravity is the structure of curved spacetime. It affects particles and light rays in that they now travel curved timelike and null geodesics in the spacetime, and it affects idealized temporal and spatial measuring instruments.

What about the second aspect of the theory, that which specified what kind of gravitational field would be generated by a source of gravity? In the older theory, any massive object generates a gravitational

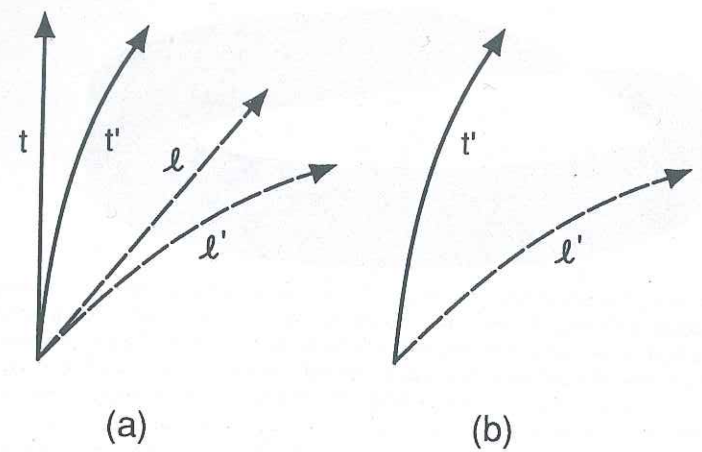

Figure 2.7 Motion in a gravitational field as following curved geodesics. In (a) spacetime is viewed as "flat." The straight line t represents the path a "free" particle would travel through the spacetime and the straight line l the path of a "free" light ray. Under the influence of a force like gravity, the particle and light ray will travel curved paths such as t' and l'. But these are viewed as deviating from the straightest paths in the spacetime. In (b) the straight paths have vanished. Instead the spacetime is viewed as "curved" in the presence of gravity, with t' and l', the paths of "free" particles and light rays (that is, particles and light rays acted upon by no nongravitational force), now considered geodesics, or straightest possible paths in the curved spacetime.

field. In the new relativistic theory, gravity is associated with the mass-energy of the material world. The field equations of general relativity have on their left-hand side a mathematical expression characterizing the curvature of spacetime. On their right-hand side they have an expression characterizing how mass-energy is distributed in spacetime, the so-called stress-energy tensor. It is this equation that relates gravity, now curved spacetime, to its sources in nongravitational mass-energy. (The "nongravitational" is important because the gravitational field itself, curved spacetime, also possesses mass-energy.) It would be a mistake to think of the matter as "causing" the gravitational field in any simplistic sense, because to know the right-hand side of the equation, which describes how mass-energy is distributed in spacetime, requires positing a spacetime structure. The equation tells us whether a given spacetime is compatible with a posited mass-energy distribution in that spacetime. Only when the equation is satisfied by both the posited spacetime structure and the posited distribution of mass-energy in that structure is the world posited a possible world in the new theory.

It is interesting that given the field equation, the dynamical law of gravity—that pointlike material particles when "free" travel timelike geodesics—follows. Unlike the Newtonian theory, the dynamical law

of gravity need not be posited as an independent law but is itself derivable from the basic field equation.

If we accept this new curved spacetime theory of gravity, we then confront the task of trying, as inhabitants of the world, to determine its actual spacetime structure. The theory tells us that the geometry of spacetime must be correlated with the distribution of matter and energy in that spacetime. And the spacetime structure in question reveals itself in terms of the curved geodesic paths of unimpeded light rays and "free" particles and by means of spatial and temporal intervals measured by ideal measuring rods (or tapes in a curved world) and clocks. Naturally, if spacetime shows curvature, it will be on astronomical scales, for we have a vast empirical experience to assure us that in local small-scale measurements, Minkowskian flat geometry works adequately.

Some effects of this newly understood "gravity as spacetime curvature" do show up in the scale of the solar system. The planets are taken to be traveling geodesics in the spacetime curved by the presence of the massive sun. This introduces slight changes from the Keplerian motion of the planets explicable by the Newtonian theory. We know that the spacetime curvature even of the solar system is small. The paths of the planets in spacetime deviate little from straight-line geodesics (not to be confused with the spatial, obviously curved, ellipses they travel). But the effect of the curvature is to superimpose on the familiar elliptical paths of the planets small additional effects, such as a motion (relative to an inertial frame fixed in the sun) of the nearest point of approach of the planet to the sun in its orbit, a motion detectable in the case of the planet Mercury.

Other metric effects of gravity can also be observed on a fairly small scale, in particular, the slowed rate of one clock relative to another if the former clock is lower down in a gravitational potential than the latter. But it is on the grand cosmological scale than the theory gives rise to its most interesting new predictions and to the possibility for the most fascinating appearances of observational consequences of spacetime curvature. Here one deals with highly idealized model universes, for which theoretical conclusions can be drawn. The hope is, of course, that at least some of these idealized pictures of the world on the cosmological scale will be close enough to reality to provide insight into the world we discover in our astronomical observations into deep space. For example, it is usually assumed that the matter of the universe can be considered as distributed uniformly and that the distribution is the same in all directions in space in the cosmological world. This assumption is now under intensive observational scrutiny.

A wide variety of possible spacetime worlds has been explored by the theorists. In many of these, the continuity structure of the world differs from that of the worlds of Newtonian or special relativistic physics. In some worlds, for example, there can be closed timelike paths, collections of events such that when an observer follows them from later to later event, he eventually returns to his initial event. Other

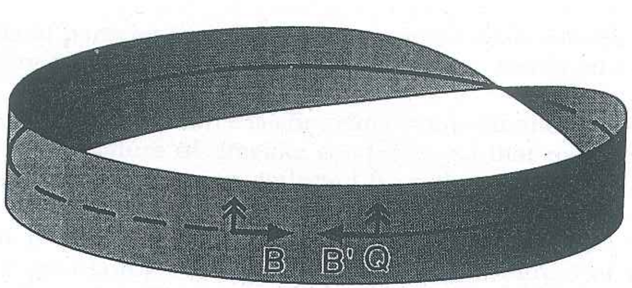

Figure 2.8 A non-orientable space. The Mobius band is the simplest example of a nonorientable space, in this case of dimension two. *B* and *B'* represent oriented figures that could not be transformed the one into the other by a rigid motion were the figures drawn on a normal plane surface. But if we take *B* and move it around the twisted Mobius band, we can eventually bring it back to *Q* so that it coincides with *B'*. This reveals the nonorientable nature of the surface. In spacetime, a nonorientability can be spatial, temporal, or spatiotemporal.

spacetimes, although not that causally pathological, can be close to having such closed, causal paths embedded in them. Other peculiar spacetimes have a nonorientability built into them. They are twisted like the familiar Mobius band, a twisted two-dimensional surface embedded in three-space. (See Figure 2.8.)

In such a nonorientable spacetime world, it may be impossible to make a global distinction between right-handed and left-handed objects, a right-handed object being transformable into a left-handed one at the same place by a voyage around the spacetime. Or there may be a lack of time orientability, which makes it impossible to say globally what is the "past" and what is the "future" direction of time at a point-event.

In some spacetimes, it is possible for observers to have spacetime split into spaces-at-a-time. That means that in these worlds, for an observer in a particular state of motion, the spacetime can be sliced into three-dimensional spaces of events that all can be assigned a specific time in a time order that can hold globally. For other spacetimes, such a splitting of spacetime into "simultaneity slices" of three-spaces-at-a-time is impossible. When such a cutting up of spacetime into spaces-at-a-time is possible, the spaces can themselves be curved three-dimensional spaces of the sort studied by Riemann in his generalization of Gauss's geometry of curved surfaces. One such universe, the Einstein model, has time going on forever in past and future. To an observer at each time the spatial world exists as a closed, three-dimensional sphere of constant, finite size. The Robertson-Walker universes have spaces-at-a-time of constant curvature, but the curvature may be positive, zero, or negative. The size parameter of these spaces can change with time, leading them to be plausible models of Big Bang universes that have, as, on the basis of the observational data, our universe seems

to have, a singular point at which all the matter of the world is compressed to one spatial point.

Moreover, spacetime curvature helps to explain the possible data of experience in another area: the description of the singularities generated by the collapsing matter of massive stars. These are the famous black holes, regions of spacetime so curved up by the presence of highly dense matter that light cannot escape to the outside spacetime from the inner spacetime region immediately surrounding the point of singular collapse of the star. Models of such locally highly curved regions of spacetime corresponding to electrically charged and/or rotating collapsed stars, as well as the original kind studied, provide fascinating case studies of the peculiar effects gravity as spacetime curvature can have. Although the evidence from observation is still inconclusive, it seems that some of the generators of highly energetic radiation in the cosmos, for example, quasars and the centers of so-called active galaxies, may very well be such black holes.

Curved Spacetime and Newtonian Gravity

When we discussed the move from space and time to spacetime when the foundations of the special theory of relativity were formulated, we noted that after Minkowski spacetime had been constructed as the spacetime appropriate to special relativity, scientists realized that one could use the spacetime notion to construct a spacetime in some ways more appropriate to the physics of Newton than was his own absolute space and time. This was the Galilean or neo-Newtonian spacetime. In light of the curved spacetime account of gravity, the general theory of relativity, it became clear that one could redescribe gravity even in the prerelativistic picture by means of a curved spacetime as well. In this prerelativistic picture, gravity doesn't have the effects on distance and time measurements it has in the relativistic version, nor is any account taken of the effect of gravity on light. Instead, it is the familiar dynamical effects of gravity that are transformed into curvature of spacetime.

In this picture, time is just as it was for Newton. There is a definite, absolute time interval between any two events. Events that are all simultaneous form spaces-at-a-time. These are, as they were for Newton, flat, three-dimensional Euclidean spaces. As in the neo-Newtonian spacetime, there is no nonrelative notion of two nonsimultaneous events being at the same place; therefore this spacetime lacks Newton's absolute notion of sameness of place through time and absolute velocity. But just as in the neo-Newtonian view there were timelike geodesics corresponding to possible paths of freely moving particles, so are there timelike geodesics in this new spacetime picture. However, whereas the timelike geodesics of the neo-Newtonian picture were the straight-line paths of uniformly moving particles (particles not acted upon by forces and, following the law of inertia, keeping their velocities constant), now, the timelike geodesics are curved lines. They are now taken to be the paths of particles "free" in the new sense made familiar from Einstein's theory of gravity, that is, acted upon by no forces other than gravity.

Once again, gravitational force is eliminated from the theory in favor of gravity as the curvature of timelike geodesics, so that particles feel the effect of gravity not by being deflected from their geodesic motion by the force of the gravitating object but, instead, by following the "free" geodesic paths in the spacetime, paths now curved as a result of the presence of the gravitating object, which serves as a "source" of spacetime curvature. Just as in Einstein's theory, it is only the uniform effect of gravity on a test object, the fact that all objects affected by gravity suffer the same modification in their motion independently of their mass or constitution, that allows for this "geometrization" of the gravitational force. This curved spacetime of Newtonian gravity is not, like Minkowski spacetime or the curved spacetime of the general theory of relativity, Riemannian (or, rather, pseudo-Riemannian) spacetime, because unlike the spacetimes of special or general relativity, it has no spacetime metric structure. There is a definite time interval between any two events. For simultaneous events, there is a definite spatial separation between any two events. In this sense, this spacetime has a metric of time and one of space. But there is, in contrast to the relativistic case, no spacetime interval between a pair of events. Curvature shows up only as the nonstraightness of the timelike geodesics, not in any metric feature of the spacetime.

Summary

So the development of the elegant theories of Einstein, which attempt to do justice to the surprising observational facts about the behavior of light, free particles, and measuring rods and clocks, provides us with two revolutions in our views of space and time. First, space and time are replaced by the unified notion of spacetime, relative to which spatial and temporal aspects of the world become derivative. Second, the notion of curvature is invoked to find a natural place for the effects of gravity in such a spacetime picture of the world.

Surely such revolutions in our scientific perspective on what space and time are actually like should result in a profound rethinking of the typically philosophical questions about space and time. How should we view the status of our claims to knowledge about the structure of space and time in this new context in which, for the first time, a variety of possible and distinct proposals about the structure of space and time are available for our scientific inspection? And what effect should such novel structures of space and time have on our views about the metaphysical nature of space and time? In particular, what effect should these revolutionary scientific views have on the traditional debate between substantivalists and relationists? It is to those philosophical questions that we now turn.

of the world sometimes tried to challenge the very logical consistency of the non-Euclidean geometries. That tack soon failed, as relative-consistency proofs for the axiomatic non-Euclidean geometries were soon produced. These proofs showed that we could be assured by pure logic that if the non-Euclidean geometries were inconsistent, then so was Euclidean geometry. Therefore, the non-Euclidean geometries were at least as logically respectable as the Euclidean. Kantians could continue to argue on other grounds that Euclidean geometry was the certain geometry of the world, holding as they did that there was a kind of necessity to the truth of Euclidean geometry that went beyond the ne-