5

Relativity: The Evolution of Einstein's Gravity

The laws of gravity are very, very strict.
And you're just bending them for your own benefit.

Billy Bragg

Icarus (Ike) Rushmore III couldn't wait to show Dieter his new Porsche. But as proud as he was of his car, he was even more excited about his Global Positioning System (GPS) that he had recently designed and installed himself.

Ike wanted to impress Dieter, so he convinced his friend to drive with him to the local track. They got in the car, Ike programmed in their destination, and the two of them set off. But to Ike's chagrin, they ended up in the wrong place—the GPS system didn't work nearly as well as he had thought it would. Dieter's first thought was that Ike must have made some ridiculous error, like confusing meters and feet. But Ike didn't believe he could have made such a stupid mistake, and he bet Dieter that wasn't the problem.

The next day, Ike and Dieter did some troubleshooting. But to their dismay, when they went for a drive the GPS was even worse than before. Ike and Dieter searched again for the problem and finally, after a frustrating week, Dieter had an epiphany. He did a quick calculation and made the startling discovery that without accounting for general relativity, Ike's GPS system would build up errors at the rate of more than 10 km each day. Ike didn't think his Porsche was fast enough to warrant relativistic calculations, but Dieter explained that the GPS signals—not the car—travel at the speed of light. Dieter modified the software to account for the changing gravitational field

L. Randall

the GPS signals had to pass through. Ike's system then worked as well as the readily available commercial variety. Relieved, Ike and Dieter began to plan a road trip.

At the beginning of the last century the British physicist Lord Kelvin said, "There is nothing new to be discovered in physics now. All that remains is more and more precise measurement."* Lord Kelvin was famously incorrect: very soon after he uttered those words, relativity and quantum mechanics revolutionized physics and blossomed into the different areas of physics that people work on today. Lord Kelvin's more profound statement, that "scientific wealth tends to accumulate according to the law of compound interest,"† is certainly true, however, and is especially appropriate to these revolutionary developments.

This chapter explores the science of gravity, and how it evolved from the impressive achievement of Newton's laws to the revolutionary advances of Einstein's theory of relativity. Newton's laws of motion are the classical physics laws that scientists used for centuries to compute mechanical motion, including motion caused by gravity. Newton's laws are magnificent, and they allow us to make predictions of motion that work spectacularly well—well enough to send men to the Moon and satellites into orbit, well enough to keep the superfast trains in Europe on the tracks when rounding corners, well enough to prompt the search for the eighth planet, Neptune, based on peculiarities in Uranus's orbit. But alas, not well enough for an accurate GPS system.

Incredibly, the GPS system now in use requires Einstein's theory of general relativity to achieve its one-meter accuracy. Determinations of the variation in snow depth on Mars using laser ranging data from orbiting spacecraft also incorporate general relativity, and yield values with an unbelievable precision of 10 cm. Certainly, at the time it was developed, no one—not even Einstein—anticipated such practical applications of a theory as abstract as general relativity.

*An address to a group of physicists at the British Association for the Advancement of Science in 1900.

†Presidential Address to British Association, 1871.

104

This chapter will explore Einstein's theory of gravity, a spectacularly accurate theory that applies to a wide range of systems. We'll begin by briefly reviewing Newton's gravitational theory, which works fine for the energies and speeds we encounter in daily life. We'll then move on to the extreme limits in which it fails: namely, very high speed (close to the speed of light) and very large mass or energy. In these limits, Newtonian gravity is superseded by Einstein's theory of relativity. With Einstein's general relativity, space (and spacetime) evolved from a static stage to a dynamical entity that can move and curve and have a rich life of its own. We'll consider this theory, the clues that led to its development, and some of the experimental tests that convince physicists that it's right.

Newtonian Gravity

Gravity is the force that keeps your feet on the ground and is the source of the acceleration that returns a tossed ball to Earth. In the late sixteenth century, Galileo showed that this acceleration is the same for all objects on the surface of the Earth, no matter what their mass.

However, this acceleration does depend on how far the object is from the Earth's center. More generally, the strength of gravity depends on the distance between the two masses—gravity's pull is weaker when objects are farther apart. And when what creates the gravitational attraction is not the Earth, but some other object, gravity's strength will depend on the mass of that object.

Isaac Newton developed the gravitational force law that summarizes how gravity depends on mass and distance. Newton's law says that the force of gravity between two masses is proportional to the mass of each of them. They could be anything: the Earth and a ball, the Sun and Jupiter, a basketball and a soccer ball, or any two objects you please. The more massive the objects, the greater the gravitational attraction.

Newton's gravitational force law also says how the gravitational force depends on the distance between the two objects. As discussed in Chapter 2, the law says that the force between two objects is

proportional to the inverse square of their separation. This inverse square law was where the famous apple entered in.* Newton could deduce the acceleration due to the Earth's gravitational pull on an apple located near the Earth's surface and compare it with the acceleration induced on the Moon, which is located sixty times further away than the Earth's surface is from its center. The acceleration of the Moon due to the earth's gravity is 3,600 times smaller (3,600 is the square of 60) than the acceleration of the apple. This is in accordance with the gravitational force decreasing as the square of the distance from the Earth's center.[7]

However, even when we know the dependence of the gravitational attraction on mass and distance, we still need another piece of information before we can determine the overall strength of gravitational attraction. The missing piece is a number, called *Newton's gravitational constant*, that factors into the calculation of any classical gravitational force. Gravity is very weak, and this is reflected in the tiny size of Newton's constant, to which all gravitational effects are proportional.

The Earth's gravitational pull or the gravitational attraction between the Sun and the planets might seem pretty big. But that's only because the Earth, the Sun, and the planets are so massive. Newton's constant is very small, and the gravitational attraction between elementary particles is an extremely weak force. This feebleness of gravity is itself a big puzzle that we will return to later on.

Although his theory was correct, Newton delayed its publication for twenty years, until 1687, while he tried to justify a critical assumption of his theory: that the Earth's gravitational pull was exerted as if its mass were all concentrated at the center. While Newton was hard at work developing calculus to solve this problem, Edmund Halley, Christopher Wren, Robert Hooke, and Newton himself made tremendous progress in determining the gravitational force law by analyzing the motion of the planets, whose orbits Johannes Kepler had measured and found to be elliptical.

These men all made major contributions to the problem of planetary motion, but it is Newton who gets credited with the inverse square

*The story might be apocryphal, but the reasoning is not.

105

law. That is because Newton ultimately showed that elliptical orbits would arise as a result of a central force (that of the Sun) only if the inverse square law was true, and he showed with calculus that the mass of a spherical body did in fact act as if it were concentrated at the center. Newton did, however, acknowledge the significance of others' contributions in his words, "If I have seen further, it is because I have stood upon the shoulders of giants."* (However, rumor has it that he said this only because of his intense dislike for Hooke, who was very short.)

In high school physics, we learned Newton's laws and calculated the behavior of interesting (if somewhat contrived) systems. I remember my outrage when our teacher, Mr Baumel, informed us that the gravitational theory we had just learned was wrong. Why teach us a theory that we know to be incorrect? In my high school view of the world, the whole merit of science was that it could be true and reliable, and could make accurate and factual predictions.

But Mr Baumel was simplifying, perhaps for dramatic effect. Newton's theory was not wrong: it was merely an approximation, one that works incredibly well in most circumstances. For a large range of parameters (speed, distance, mass, and so on), it predicts gravitational forces quite accurately. The more precise underlying theory is relativity, and you only make measurably different predictions with relativity when you are dealing with extremely high speeds or large amounts of mass or energy. Newton's law predicts the motion of a ball admirably well, since neither of the above criteria apply. To use relativity to predict the motion of a ball would be pure silliness.

In fact, Einstein himself initially thought of special relativity merely as an improvement on Newtonian physics—not as a radical paradigm shift. This, of course, grossly underplays the ultimate significance of his work.

*Letter from Isaac Newton to Robert Hooke, 5 February 1675.

Special Relativity

A very reasonable thing to expect from physical laws is that they should be the same for everyone. No one could blame us for questioning their validity and utility if people in different countries or sitting on moving trains or flying on an airplane experienced different physical laws. Physical laws should be fundamental and hold true for any observer. Any differences in calculations should be due to differences in environment, not the physical laws. It would be very strange indeed to have universal physical laws that required a particular vantage point. The particular quantities you might measure could depend on your reference frame, but the laws that govern these quantities should not. Einstein's formulation of special relativity ensures that this is the case.

In fact, it's somewhat ironic that Einstein's work on gravity is referred to as "the theory of relativity." The essential point that drove both special and general relativity was that physical laws should apply for everyone, independent of their reference frame. In fact, Einstein would have preferred the term *Invariantentheorie*.* In a letter Einstein wrote in 1921 in reply to a correspondent who had suggested he reconsider the name, he admitted that the term "relativity" was unfortunate.† But by that time, the term was too well entrenched for him to attempt to change it.

Einstein's first insight about reference frames and relativity came from thinking about electromagnetism. The well-known theory of electromagnetism from the nineteenth century was based on Maxwell's laws, which describe the behavior of electromagnetism and electromagnetic waves. The laws gave correct results, but everyone initially falsely interpreted the predictions in terms of the motion of an *aether*, a hypothesized invisible substance whose vibrations were supposed to be electromagnetic waves. Einstein realized that if there were an aether, there would also be a preferred observational vantage

*Gerald Holton, *Einstein, History, and Other Passions* (Cambridge, MA: Harvard University Press, 2000).
†Letter to E. Zschimmer, 30 September 1921.

106

point, or frame of reference: the one in which the aether is at rest. He reasoned that the same physical laws should apply to people who are moving at constant velocity* with respect to each other and with respect to someone at rest—that is, in frames of reference that physicists refer to as inertial frames. By requiring that *all* physical laws, including those of electromagnetism, should hold for observers in all inertial reference frames, Einstein was led to abandon the idea of the aether and, ultimately, to formulate special relativity.

Einstein's theory of special relativity, with its radical revision of the concepts of space and time, was a major leap. Peter Galison,† a physicist and historian of science, suggests that it was not only the aether theory that put Einstein on the right track, but Einstein's job at the time. Galison reasoned that Einstein, who grew up in Germany and worked at the patent office in Bern, Switzerland, must have had time and time coordination on his mind. Anyone who has traveled in Europe knows that precision is valued highly in countries such as Switzerland and Germany, which has the happy consequence that passengers can count on the trains to run on time. Einstein worked in the patent office between 1902 and 1905, during an era when train travel was becoming increasingly important, and coordinating time was at the forefront of new technology. In the early 1900s, Einstein was very likely thinking about real-world problems, such as how to coordinate the time at one train station with that at another.

Of course, Einstein did not need to develop relativity to solve the problem of coordinating real trains. (For those of us accustomed to the frequently delayed American trains, coordinated time might sound exotic in any case.‡) But coordinating time raised some interesting questions. Time coordination of relativistically moving trains is not a straightforward problem. If I were to coordinate my watch with someone on a moving train, I would need to account for the time delay of a signal traveling between us because light has a finite speed. Coordinating my watch with that of the person sitting next to me

*Velocity gives both speed and direction.

†Peter Galison, *Einstein's Clocks, Poincaré's Maps: Empires of Time* (New York: W.W. Norton, 2003).

‡Don't get me wrong—I like trains. But I wish they were better supported in the U.S.

would not be the same as coordinating watches with someone far away.*

Einstein's critical insight, the one that led him to special relativity, was that ideas about time had to be reformulated. According to Einstein, space and time could no longer be considered independently. Although they are not the same thing—time and space are clearly different—the quantities you measure depend on the speed at which you are traveling. Special relativity was the result of this insight.

Bizarre as they are, one can derive all of Einstein's novel consequences of special relativity from two postulates. To state them, we need to understand the meaning of *inertial frames*—a particular category of reference frames. Let's first choose any frame of reference that moves at constant velocity (speed and direction); the one that's at rest is often a good one. The inertial frames would then be those that are moving at fixed velocity with respect to that first one—someone running or driving by at constant speed, for example.

Einstein's postulates then state that:

> The laws of physics are the same in all inertial frames.
> The speed of light, *c*, is the same in any inertial frame.

The two postulates tell us that Newton's laws are incomplete. Once we accept Einstein's postulates, we have no choice but to replace Newton's laws with new physical laws that are consistent with these rules.[8] The laws of special relativity that follow lead to all the surprising consequences you might have heard of, such as time dilation, the observer dependence of simultaneity, and Lorentz contraction of a moving object. The new laws should look very much like the old classical physics laws when applied to objects moving at speeds that are small compared with the speed of light. But when applied to something moving very fast, at or near the speed of light, the difference

*Although American trains don't always coordinate time very well, Amtrak does appear to acknowledge special relativity when they say, "time and the space.to use it" in their advertising slogan for the Acela, the high-speed train that travels the Northeast corridor. However, "time" and "space" are not precisely interchangeable. Although the slogan "space and the time to use it" does describe my more heavily delayed train rides, the phrase wouldn't be a very compelling advertisement for a high-speed train.

between the Newtonian and special relativity formulations should become apparent.

For example, in Newtonian mechanics speeds are simply added together. A car driving towards yours on the freeway approaches you at a speed that's the sum of its speed and yours. Similarly, if someone throws a ball at you from the platform while you are on a moving train, the ball's speed appears to be the sum of the speed of the ball itself plus the speed of the moving train. (A former student of mine, Witek Skiba, can attest to this fact. Witek was nearly knocked out when he was hit by a ball that someone threw at the approaching train he was riding.)

According to Newtonian physics, the speed of a beam of light directed at a moving train should be the sum of the speed of light and the speed of the moving train. But this can't be true if the speed of light is constant, as Einstein's second postulate asserts. If the speed of light is always the same, then the speed of the beam aimed at the moving train will be identical to the speed of a light beam that approaches you when you're standing still on the ground. Even though it runs counter to the intuition gained from your experience of the slow speeds you encounter in daily life, light speed is constant, and in special relativity speeds don't simply add up as they do in Newtonian physics. Instead, you add speeds according to a relativistic formula that follows from Einstein's postulates.

Many of special relativity's implications don't jibe with our familiar notions of time and space. Special relativity treats time and space differently than they had been treated before in Newtonian mechanics, and this is what gives rise to many of its counterintuitive results. Time and space measurements depend on speed and get mixed up in systems that move relative to each other. Nonetheless, surprising as they are, once you accept the two postulates then a different notion of space and time is an inevitable consequence.

Here's one argument why. Imagine two identical ships with identical masts. One ship is docked by the shore, while the other is moving away. Also imagine that the captains of the two ships synchronized their watches when the first ship sailed off.

Now suppose that the two captains do a rather odd thing: each decides to measure time on her ship by placing a mirror at the top of

the mast and a second mirror at the bottom, shining a light from the bottom mirror to the top one, and measuring the number of times light hits the mirror and returns. As a practical matter, of course, this would be absurd, since light would cycle up and down far too frequently to count. But bear with me, and imagine that the captains can count extraordinarily fast; I'll be using this somewhat contrived example to argue that time stretches out on the moving ship.

If each captain knows how long it takes for light to cycle once, she can calculate the passage of time by multiplying the light-cycle time by the number of times light cycles up and down between the mirrors. Now suppose, though, that instead of using her own stationary mirror clock, the captain on the docked ship measures time by the number of times the light on the moving ship hits the mast's mirror and returns.

Now from the perspective of the captain on the moving ship, the light simply goes straight up and down. However, from the perspective of the captain on the docked ship, the light has to travel farther (in order to cover the distance traveled by the moving ship—see Figure 35). But—and this is the counterintuitive part—the speed of light is constant. It is the same for the light sent to the top of the mast on the docked ship as it is for the light sent to the top of the mast on the moving ship. Since speed measures distance traveled over time, and the speed of light for the moving ship is the same as the speed of light for the stationary one, the moving mirror clock has to "tick" at a slower rate to compensate for the longer distance the moving light

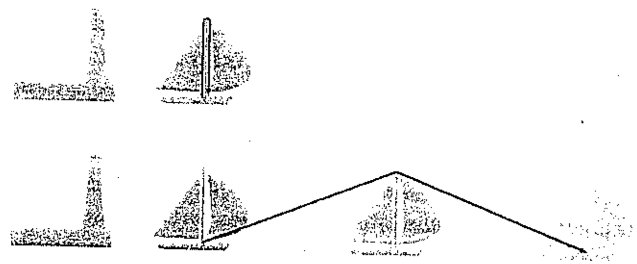

Figure 35. *The path of a light beam that bounces off the top of a mast of a stationary ship and of a moving one. The stationary observer (in a boat by the shore or in a lighthouse) would see a longer path in the second case.*

108

has to travel. This very counterintuitive conclusion—that moving and stationary clocks must tick at different rates—follows from the fact that the speed of light in a moving reference frame is the same as the speed of light in a stationary one. And although this is a funny way to measure time, the same conclusion—that moving clocks run slower—would hold true independently of how time is measured. If the captains had watches on, they would observe the same thing (again, with the caveat that for normal speeds, the effect would be tiny).

While the above example is artificial, the phenomenon described produces genuinely measurable effects. For example, special relativity gives rise to the different time experienced by fast-moving objects—the phenomenon known as time dilation.

Physicists measure time dilation when they study elementary particles produced at colliders or in the atmosphere, which travel at relativistic speeds—speeds approaching that of light. For example, the elementary particle called a muon has the same charge as an electron, but is heavier and can decay (that is, it can turn into other, lighter particles). The muon's lifetime, the time before it decays, is only 2 microseconds. If a moving muon had the same lifetime as a stationary one, it would be able to travel only about 600 meters before it disappeared. But muons manage to make it all the way through our atmosphere, and in colliders, to the edges of large detectors, because their near-light-speed velocity makes them appear to us much longer-lived. In the atmosphere, muons travel at least ten times further than they would in a universe based on Newtonian principles. The very fact that we see these muons at all shows us that time dilation (and special relativity) gives rise to true physical effects.

Special relativity is important both because it was a dramatic deviation from classical physics and because it was essential to the development of general relativity and quantum field theory, both of which play a significant role in more recent developments. Because I won't use specific special relativity predictions when I discuss particle physics and extra-dimensional models later on, I'll resist the urge to go into all the fascinating consequences of special relativity, such as why simultaneity depends on whether an observer is moving and how the sizes of moving objects are different from when they are at rest.

Instead, we'll delve into another dramatic development, namely general relativity, which will be critical when we consider string theory and extra dimensions later on.

The Principle of Equivalence: General Relativity Begins

Einstein wrote down his theory of special relativity in 1905. In 1907, while working on a paper that summarized his recent work on the subject, he found himself already questioning whether the theory could apply to all situations. He noticed two major omissions. For one thing, physical laws looked the same only in certain special inertial reference frames—those that moved at fixed velocity with respect to each other.

In special relativity, these inertial reference frames occupied a privileged position. The theory left out any reference frame that was accelerating. If you pressed the accelerator while driving your car, you would no longer be in one of the special reference frames where the laws of special relativity apply. That's what's "special" in special relativity: the "special" inertial frames are only a small subset of all possible reference frames. For someone convinced that no one's reference frame is special, it was a big problem that the theory singled out inertial reference frames.

Einstein's second misgiving concerned gravity. Although he had figured out how objects respond to gravity in some situations, he still hadn't come up with formulas for determining the gravitational field in the first place. The form of the gravitational force law was known in some simpler settings, but Einstein wasn't yet able to deduce the field for every possible distribution of matter.

Between 1905 and 1915, in a sometimes grueling exploration, Einstein addressed these problems. The result was general relativity. He centered his new theory around the *equivalence principle*, which states that the effects of acceleration cannot be distinguished from those of gravity. All the laws of physics would look the same to an accelerating observer as they would to a stationary observer placed in a gravitational field that accelerates everything in the stationary frame

109

with an acceleration of the same magnitude—but in the opposite direction—as the original observer's acceleration. In other words, you wouldn't have any way of distinguishing uniform acceleration from standing still in a gravitational field. According to the principle of equivalence, there is no measurement that would distinguish between these two situations. An observer could never know which situation he was in.

The equivalence principle follows from the equivalence of *inertial* and *gravitational mass*, two quantities that in principle could have been different from each other. Inertial mass determines how an object will respond to any force—how much the object would accelerate if you applied that force. The role of inertial mass is summarized in Newton's second law of motion, $F = ma$, which says that if you apply a force of magnitude F to an object with mass m, you will produce an acceleration a. Newton's famous second law tells us that a given force produces smaller acceleration on an object that has bigger inertial mass, which is probably very familiar to you from experience. (If you shove a footstool, it will go further and faster than if you shove a grand piano just as hard.) Notice that this law applies for any sort of force—the force of electromagnetism, for example. It can apply in situations that have nothing whatsoever to do with gravity.

Gravitational mass, on the other hand, is the mass that enters the gravitational force law and determines the strength of gravitational attraction. As we saw, the strength of the Newtonian gravitational force is proportional to the two masses that get attracted to each other. These masses are gravitational mass. Gravitational mass and the inertial mass that enters Newton's second force law turn out to be the same, and that's why we can safely give them the same name: mass. But in principle they could have been different, and we would have had to call one "mass" and the other "ssam." Fortunately, we don't need to do that.

The mysterious fact that the two masses are the same has deep implications, which it took an Einstein to recognize and develop. The gravitational force law states that the strength of gravity is proportional to mass, and Newton's law tells us how much acceleration would be generated by that (or any other) force. Because the strength of gravity is proportional to the same mass that determines the amount

of acceleration, the two laws together tell us that even though the *force* depends on mass through $F = ma$, the acceleration induced by gravity is entirely independent of the mass that gets accelerated.

The acceleration of gravity that any object experiences must be the same for anyone or anything separated by the same distance from another object. This is the claim that Galileo allegedly verified by dropping objects off the Tower of Pisa,* demonstrating that the Earth induces the same acceleration for all objects, independent of their mass. This fact—that acceleration is independent of the mass of the accelerated object—is unique to the gravitational force, because the strength of no force other than gravity depends on mass. And because the gravitational force law and Newton's law of motion depend on mass in the same way, the mass cancels out when you calculate acceleration. Acceleration therefore doesn't depend on mass.

This relatively straightforward deduction has profound implications. Since all objects have the same acceleration in a uniform gravitational field, if this *single* acceleration could be canceled, the evidence of gravity would be canceled as well. And that is exactly what happens to a freely falling body: it is accelerated precisely so as to cancel the evidence of gravity.

The equivalence principle says that if you and everything around you were freely falling, you would not be aware of a gravitational field. Your acceleration would cancel the acceleration that the gravitational field would otherwise have produced. This state of weightlessness is now familiar from pictures from orbiting spacecraft, where the astronauts and the objects that surround them don't experience any gravity.

Textbooks often illustrate the absence of gravity's effects (from the vantage point of the freely falling observer) with a picture of someone dropping a ball in a free-falling elevator. You see the person and the ball falling together in the picture. The person in the elevator would always see the ball at the same height above the elevator's floor. He wouldn't see the ball drop (see Figure 36).

Physics texts always present the freely falling elevator as if it were the most natural thing in the world that the observer inside would

*He did the experiment by timing objects rolling down an inclined plane.

110

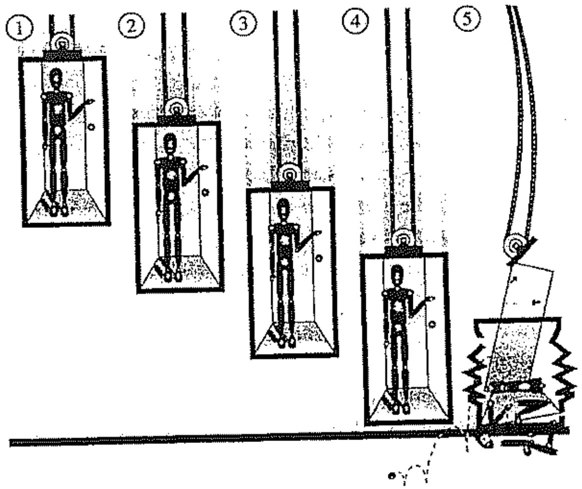

Figure 36. *An observer in a falling elevator who releases a ball will not see it drop. However, when a freely falling elevator meets the stationary Earth, the observer will not be very happy.*

calmly watch a ball not drop with complete equanimity, with no concern at all for his personal well-being. This is in sharp contrast to the terrified faces in movies in which the cables of an elevator are cut and the actors hurtle towards the ground. Why such different responses? If everything were freely falling, there would be no cause for alarm. The situation would be indistinguishable from everything being at rest, albeit in a zero-gravity environment. But if, as in the movies, someone is falling but the ground below him stays put, he has good reason to be petrified. If someone is on a freely falling elevator, but solid ground awaits his descent, you can be sure that he will notice the consequences of gravity when his free fall is ended (as is illustrated in the last frame of Figure 36).

The reason that Einstein's conclusion seems so surprising and strange is that our upbringing here on Earth, with a stationary planet beneath our feet, biases our intuition. When the force of the Earth keeps you stationary on the ground, you notice the effects of gravity because you are not following the path towards the center of the Earth

that gravity would have you follow. On Earth, we're accustomed to gravity making things fall. But "falling" really means "falling relative to us." If we were falling along with a dropped ball, as we would be in a free-falling elevator, the ball would not go down any faster than we would. We therefore would not see it drop.

In your freely falling reference frame, all the laws of physics would coincide with the laws of physics that would be obeyed if you and everything near you were at rest. A freely falling observer would observe that motion is described by the same equations, consistent with special relativity, that apply for observer in an inertial, non-accelerating reference frame. In the review paper he wrote in 1907 about relativity, Einstein explains how the gravitational field has only a relative existence, "because for an observer falling freely from the roof of a house there exists—at least in his immediate surroundings—no gravitational field."*

This was Einstein's major insight. The equations of motion for a freely falling observer are the equations of motion for an observer in an inertial reference frame. A freely falling observer does not feel the force of gravity—only objects that are not in free fall experience a gravitational force.

In our lives we don't generally encounter things or people in free fall. When free fall happens, it looks scary and dangerous. But, as an Irishman said to the physicist Raphael Bousso when he was visiting Ireland's Cliffs of Moher, "It's not the fall that kills you, but the %&!# crash when you stop." And when I broke several bones in a rock-climbing accident and had to miss a conference I had organized, there were quite a few jokes about my testing the theory of gravity. I can state with complete confidence that gravitational acceleration agrees with predictions.

*Albert Einstein, "Über das Relativitätsprinzip und die aus demselben gezogene Folgerungen" ["On the relativity principle and the conclusions drawn from it"], *Jahrbuch der Radioaktivität und Electronik*, vol. 4, pp. 411–62 (1907); see also Abraham Pais, *Subtle is the Lord* (Philadelphia: American Philological Association, 1982).

Tests of General Relativity

There's more to general relativity; soon we'll get to the rest, which took considerably longer to develop. But the equivalence principle alone explains many results from general relativity. Once Einstein had recognized that gravity could be canceled in an accelerating reference frame, he could calculate gravitational influence by imagining an accelerating system equivalent to the one with gravity. This allowed him to calculate the gravitational effects for some interesting systems which others could use to check his conclusions. We'll now consider a few of the most significant experimental tests.

First is the *gravitational redshift* of light. A redshift causes us to detect light waves at a lower frequency than the frequency at which they were emitted. (You've probably encountered the analogous effect in sounds waves when a motorcycle roared past you and the sound waves rose and fell in pitch.)

There are several ways to understand the origin of the gravitational redshift, but probably the simplest is through an analogy. Imagine that you throw a ball up into the air. The rising ball slows down as it moves against the force of gravity. But the ball's energy is not lost, even though the ball is slowing down. It is converted into potential energy, which is then released as kinetic energy, or energy of motion, when the ball falls back down.

The same reasoning applies to the particle of light, the *photon*. Just as a ball loses momentum when it is thrown up into the air, a photon loses momentum as it escapes from a gravitational field. As with the ball, this means that the photon loses kinetic energy but gains potential energy as it fights its way out of the gravitational field. But a photon cannot slow down as a ball would, since it always travels at the constant speed of light. To jump the gun a bit, we will see in the next chapter that one consequence of quantum mechanics is that a photon lowers its energy when it lowers its frequency. And that is exactly what happens to the photon that is going through the changing gravitational potential. In order to lower its energy, the photon decreases its frequency, and this lowered frequency is the gravitational redshift.

Conversely, a photon that was moving towards a gravitational source would increase its frequency. In 1965, the Canadian-born physicist Robert Pound and one of his students, Glen Rebka, measured this effect by studying gamma rays emitted from radioactive iron that was placed at the top of the "tower" of Harvard's Jefferson Lab, the building where I now work. (Though it's part of the building, an elevated attic area in Jefferson Lab and the floors beneath it are known as "the tower."). The gravitational fields at the top and bottom of the tower were slightly different, since the top is slightly further from the center of the Earth. A high tower would be best for this measurement, since it would maximize the difference in height between where the gamma rays were emitted (the top of the tower) and where they were detected (the basement). But even though the tower consists of just three floors, an attic, and some windows that peer out above the attic—it's all of 74 feet high—Pound and Rebka managed to measure the difference in frequency between the emitted and absorbed photons with incredible precision, five parts in a million billion. They thereby established that the general relativity predictions for the gravitational redshift were correct to at least 1% accuracy.

A second experimentally observable consequence of the equivalence principle is the bending of light. Gravity can attract energy as well as mass. After all, the famous relation $E = mc^2$ means that energy and mass are closely connected. If mass experiences gravity, then so should energy. The Sun's gravity influences mass, and likewise affects the trajectory of light. Einstein's theory predicted exactly the amount light should bend under the Sun's influence. These predictions were first confirmed during the solar eclipse of 1919.

The English scientist Arthur Eddington organized expeditions to the island of Principe off the coast of West Africa and to Sobral in Brazil, where the eclipse could best be seen. Their purpose was to photograph the stars in the neighborhood of the eclipsed Sun and check whether stars that appeared near the Sun moved relative to their usual positions. If the stars did appear to be shifted, that would mean that their light was traveling along a bent trajectory. (The scientists needed to make their measurements during an eclipse so that the sunlight wouldn't overwhelm the much dimmer light of the stars.) Sure enough, the stars appeared in just the right "wrong" places. The

112

measurement of the correct bending angle provided strong evidence supporting Einstein's theory of general relativity.

Incredibly, the bending of light is now so well established and understood that it is one of the tools that was used to probe the distribution of mass in the universe and look for dark matter in the form of small, burnt-out stars that no longer emit light. Like black cats on a moonless night, such objects are very hard to see. The only way to observe them is through their gravitational effects.

Gravitational lensing is one way that astronomers can learn about dark objects; dark objects, like everything else, interact via gravity. Although the burnt-out stars do not themselves emit light, there can be bright objects behind them (from our perspective) whose light we can see. Without any dark star near its path, the light would travel in straight lines. But light emitted by a bright star will bend when it passes by the dark star. Light passing on the left will bend in the opposite direction than light passing on the right and light passing on the top will bend in the opposite direction than light passing on the bottom. This will create multiple images of a bright object behind a dark star and the effect is called *gravitational lensing*. Figure 37 shows an example of a multiple image of a quasar that appeared when an intervening massive object bent the quasar's light rays in different directions.

The Graceful Curves of the Universe

The equivalence principle says that the force of gravity is indistinguishable from constant acceleration. I'm glad you made it to this point, because I need to confess that I simplified, and the two aren't entirely indistinguishable after all. How could they be? If gravity were equivalent to acceleration, it would not be possible for people in opposite hemispheres to simultaneously fall to Earth. After all, the Earth cannot accelerate in two directions at once. Gravitational pull in the different directions felt in America and China, for example, cannot possibly be accounted for by a single acceleration.

The resolution of this paradox is that the equivalence principle asserts only that gravity can be replaced by acceleration *locally*. At

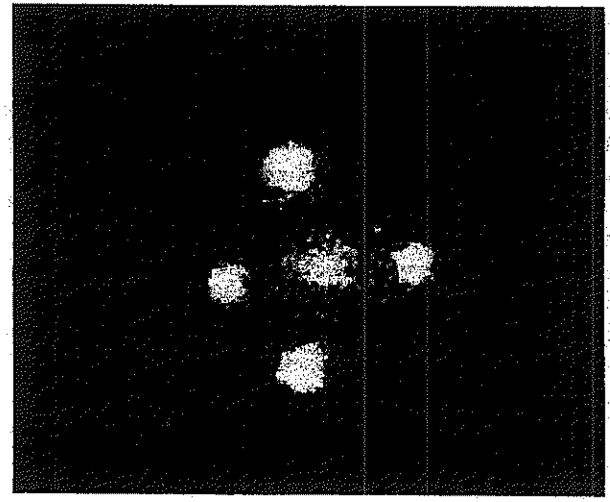

Figure 37. *The "Einstein Cross" is formed when multiple images of a bright, distant quasar are formed by light bending in different directions as it passes by a massive foreground galaxy.*

different places in space, the acceleration that would replace gravity according to the principle would generally be in different directions. The answer to our problem with Chinese/American relations is that American gravity is equivalent to an acceleration in a different direction from the acceleration that would reproduce Chinese gravity.

This critical insight led Einstein to a complete reformulation of the theory of gravity. He no longer saw gravity as a force that acts directly on an object. Instead, he described it as a distortion of the geometry of spacetime that reflects the different accelerations required to cancel gravity in different places. Spacetime is no longer a parenthetical background to an event—it is an active player. With Einstein's theory of general relativity, the force of gravity is understood in terms of the curvature of spacetime, which in turn is determined by the matter and energy that are present. Let's now consider the notion of the curvature of spacetime, on which Einstein's revolutionary theory rests.

Curved Space and Curved Spacetime

A mathematical theory must be internally consistent but, unlike a scientific theory, it has no obligation to correspond to an external physical reality. True, mathematicians have often drawn inspiration from what they see in the world around them. Mathematical objects such as cubes and natural numbers do have real-world counterparts. But mathematicians extend their assumptions about these familiar concepts to objects whose physical reality is less certain, such as tesseracts (hypercubes in four-dimensional space) and quaternions (an exotic number system).

Euclid wrote his five fundamental postulates of geometry in the third century BC. From these assumptions a beautiful logical structure developed, one that you might have had a taste of in high school. But later mathematicians found themselves having trouble with the fifth postulate, the one known as the parallel postulate. This postulate states that, given a line and a point outside that line, there is one and only one line that can be drawn through the point that is parallel to the initial line.

For two millennia after Euclid formulated his postulates, mathematicians argued about whether this fifth postulate was actually independent or merely a logical consequence of the other four. Could there be a system of geometry for which all but the last postulate was true? If no such system of geometry existed, the fifth postulate would not be independent, and would therefore be disposable.

Only in the nineteenth century did mathematicians put the fifth postulate in its proper place. The great German mathematician Carl Friedrich Gauss discovered that Euclid's fifth assumption was exactly what Euclid had claimed: a postulate that could be replaced by another. He went ahead and replaced it, discovering other systems of geometry and thereby demonstrating that the fifth postulate was independent. With that, non-Euclidean geometry was born.

A Russian mathematician, Nikolai Ivanovich Lobachevsky, also developed non-Euclidean geometry, but when he sent his work to Gauss he was disappointed to learn that the older mathematician had come up with the same idea fifty years before. But neither Lobachevsky

nor anyone else had known about Gauss's results, which the German had hidden for fear that his colleagues would ridicule him.

Gauss shouldn't have worried. It is obvious that Euclid's fifth postulate is not always true, because we all know about alternatives. For example, lines of longitude meet at the North Pole and at the South Pole, even though they are parallel at the equator. Geometry on a sphere is an example of non-Euclidean geometry. Had the ancients written on spheres rather than scrolls, this might have been obvious to them, too.

But there are many examples of non-Euclidean geometry which, unlike the sphere, cannot be realized physically in a three-dimensional world. The original non-Euclidean geometries of Gauss, Lobachevsky, and the Hungarian mathematician János Bolyai* dealt with such undrawable theories, which makes it less surprising that they took so long to discover them.

A few examples illustrate what makes curved geometries different from the flat geometry of this page. Figure 38 shows three two-dimensional surfaces. The first, the surface of a sphere, has constant positive curvature. The second, a section of a flat plane, has zero

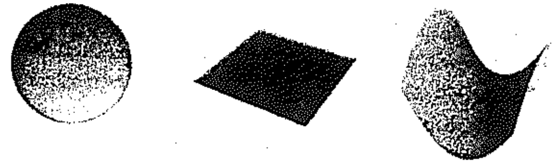

Figure 38. *Surfaces of positive, zero, and negative curvature.*

*János Bolyai was a genius, but although his father, Farkas Bolyai, wanted him to be a mathematician, János was poor and joined the military and not the academy. Others initially discouraged János about his work on non-Euclidean geometry, and he eventually published it only because his father insisted on putting it in a book he was writing. Farkas, who was friends with Gauss, sent him the appendix that János wrote. But once again, János was in for disappointment. Although Gauss recognized János Bolyai's genius, he replied only, "To praise it would amount to praising myself. For the entire content of the work . . . coincides almost exactly with my own meditations which have occupied my mind for the past thirty or thirty-five years." (Letter from Gauss to Fartas Bolyair, 1832.) So once again, János's mathematical career was thwarted.

114

curvature. And the third, a hyperbolic paraboloid, has constant negative curvature. Examples of negatively curved surfaces are the shape of a horse's saddle, the terrain between two mountain peaks, and a Pringles potato chip.

There are many litmus tests that will tell us which of the three possible types of curvature any particular geometric space possesses. For example, you can draw a triangle on each of the three surfaces. On the flat surface the sum of the angles of a triangle is always precisely 180 degrees. But what about a triangle on the surface of the sphere, with one vertex on the North Pole and the remaining two vertices on the equator, a quarter of the way around the equator from each other? Each of the angles of this triangle is a right angle of 90 degrees. Therefore the sum of the angles on the triangle is 270 degrees. This could never happen on a flat surface, but on a surface of positive curvature the sum of the angles of a triangle must exceed 180 degrees because the surface bulges out.

Similarly, the sum of the angles of a triangle drawn on a hyperbolic paraboloid is always less than 180 degrees, a reflection of its negative curvature. This is a bit harder to see. Draw two vertices near the top of the saddle and one down low, along one of the lower parts of the hyperbolic paraboloid, where one of your feet would go if you were sitting on a horse. This last angle is less than it would be if the surface were flat. The angles add up to less than 180 degrees.

Once it was established that non-Euclidean geometries were internally consistent—that is, their premises didn't result in paradoxes or contradictions—the German mathematician Georg Friedrich Bernhard Riemann developed a rich mathematical structure to describe them. A piece of paper cannot be rolled into a sphere, but it can be rolled into a cylinder. You can't flatten a saddle without having it crumple or fold back on itself. Building on Gauss's work, Riemann created a mathematical formalism that encompassed such facts. In 1854 he found a general solution to the problem of how to characterize all geometries through their intrinsic properties. His studies laid the groundwork for the modern mathematical field of differential geometry, a branch of mathematics that studies surfaces and geometry.

Because I will almost always consider space and time together from now on, we will generally find the notion of *spacetime* more useful than the notion of space. Spacetime has one more dimension than space: in addition to "up-down," "left-right," and "forwards-backwards," it includes time. In 1908 the mathematician Hermann Minkowski used geometric notions to develop this idea of an absolute spacetime fabric. Whereas Einstein studied spacetime using time and space coordinates that depended on a frame of reference, Minkowski identified the observer-independent spacetime fabric that can be used to characterize a given physical situation.

In the rest of the book, when I refer to dimensionality I will be giving the number of spacetime dimensions, except where I explicitly state otherwise. For example, when we look around us we see what I will from now on refer to as a four-dimensional universe. Occasionally I will single out time and talk about a "three-plus-one"-dimensional universe, or three spatial dimensions. Bear in mind that all these terms refer to the same setting—one that has three dimensions of space and one of time.

The spacetime fabric is a very important notion. It concisely characterizes the geometry that corresponds to the gravitational field produced by a particular distribution of energy and matter. But Einstein initially disliked the idea, which had seemed to him like an overly fancy way to reformulate the physics that he had already explained. However, he eventually recognized that the spacetime fabric was essential for completely describing general relativity and calculating gravitational fields. (For the record, Minkowski wasn't overly impressed with Einstein on first acquaintance, either. Based on Einstein's performance in Minkowski's calculus class back when Einstein was a student, Minkowski had concluded then that Einstein was a "lazy dog.")

Einstein wasn't alone in resisting non-Euclidean geometry. His friend Marcel Grossmann, a Swiss mathematician, also considered it unduly complicated and tried to talk Einstein out of using it. However, they eventually agreed that the only tractable way of explaining gravity was by using non-Euclidean geometry to represent the spacetime fabric. Only then could Einstein interpret and calculate the warping of spacetime that was equivalent to gravity, which turned out to be the key to completing general relativity. After Grossmann conceded defeat, both he and Einstein struggled through the intricacies of differential geometry to simplify their highly complicated earlier attempts to arrive at a formulation of the theory of gravity. In the end, they completed the theory of general relativity and reached a deeper understanding of gravity itself.

Einstein's Theory of General Relativity

General relativity presented a radical revision of the concept of gravity. We now understand gravity—the force that keeps your feet on the ground and binds together our galaxy and the universe—not as a force acting directly on objects, but as a consequence of the geometry of spacetime, an idea that took Einstein's view of the union of space and time to its logical conclusion. General relativity exploits the deep connection between inertial and gravitational mass to formulate the effect of gravity *solely* in terms of the geometry of spacetime. Any distribution of matter or energy curves or warps spacetime. Curved pathways in spacetime determine gravitational motion, and the matter and energy of the universe cause spacetime itself to expand, undulate, or contract.

In flat space the shortest distance between two points, the *geodesic,* is a straight line. In curved space we still can define a geodesic as the shortest path between two points, but that path won't necessarily look straight. For example, routes of airplanes that follow great circles on the Earth are geodesics. (A great circle is any circle, such as the equator or a line of longitude, that goes around the fattest part of a sphere.) Although these paths are not straight, they are the shortest routes that don't tunnel through the Earth.

In curved four-dimensional spacetime, we can also define a geodesic. For two events separated in time, a geodesic is the natural path things would take in spacetime to connect one event to the other. Einstein realized that free fall, which is the path of least resistance, is motion along the spacetime geodesic. He concluded that, in the absence of external forces, dropped objects will fall along a geodesic, as with the

path of the person on the falling elevator who doesn't feel his weight or see a ball drop.

However, even when things are following their geodesics through spacetime and there are no external forces, gravity has noticeable effects. We've already seen that the local equivalence between gravity and acceleration was one of the critical insights that led Einstein to develop an entirely new way of thinking about gravity. He deduced that, because the acceleration induced by a gravitational force is locally the same for all masses, gravity must be a property of spacetime itself. That's because "freely falling" means different things in different places, and it is only *locally* that gravity can be replaced by a single acceleration. My Chinese counterpart and I fall in different directions, even if we are both in our local version of Einstein's elevator. The fact that the direction of free fall is not the same everywhere is a reflection of the curvature of spacetime. There isn't a *single* acceleration that can cancel the effects of gravity everywhere. In curved spacetime, the geodesics of different observers will in general be different. So globally, gravity has observable consequences.

General relativity goes much further than Newtonian gravity because it allows us to calculate the relativistic gravitational field of any distribution of energy and matter. Moreover, the revelation that the geometry of spacetime encodes the effects of gravity permitted Einstein to close a major gap in his original formulation of gravity. Although physicists at the time knew how objects would respond to a gravitational field, they did not know what gravity was. Now they understood that the gravitational field is the distortion of the spacetime fabric caused by matter and energy. This distortion extends throughout the cosmos itself, or, as we will see shortly, throughout a higher-dimensional spacetime that might include branes. All of the gravitational effects of these more complicated situations can be embedded in the ripples and curves of a spacetime surface.

A picture gives perhaps the best description of how matter and energy distort the spacetime fabric to create a gravitational field. Figure 39 shows a sphere of matter sitting in space. The space surrounding the sphere is distorted: the ball makes a depression in the spatial surface whose depth reflects the ball's mass or energy. A ball

116

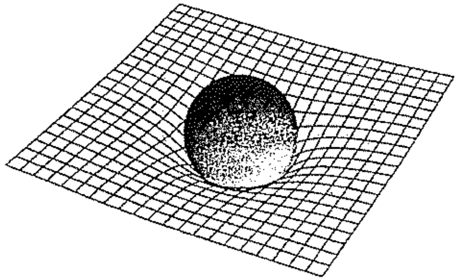

Figure 39. *A massive object distorts the surrounding space, thereby creating a gravitational field.*

passing nearby will roll towards the central depression, where the mass is located. According to general relativity, the spacetime fabric warps in an analogous fashion. Another ball passing through would be accelerated towards the center of the sphere. In this case, the result would agree with what Newton's law would predict, but the interpretation and calculation of the motion would be very different. According to general relativity, a ball follows the undulations of the spacetime surface, and thereby implements the motion induced by the gravitational field.

Figure 39 is a bit misleading, so you should keep in mind several caveats. First of all, I've shown the space surrounding the ball as two-dimensional. But really, the full three-dimensional space and the full four-dimensional spacetime are warped. Time is warped because it too is a dimension from the vantage point of special and general relativity. Warped time is how special relativity tells us that clocks run at different rates in different places, for example. A further caveat is that a second ball rolling in the curved geometry around the first ball would also affect the geometry of spacetime; we have assumed that its mass is much smaller than the larger ball's and neglected this small effect. The third thing that's important to keep in mind is that the object distorting spacetime can have any number of dimensions. Later on, a brane will play the role of the sphere in this picture.

Nonetheless, in all cases matter tells spacetime how to curve, and spacetime tells matter how to move. Curved spacetime sets up the

geodesic paths along which, in the absence of other forces, things will travel. Gravity is encoded into the geometry of spacetime. It took Einstein the better part of a decade to deduce this precise connection between spacetime and gravity, and to incorporate the effects of the gravitational field itself—after all, the gravitational field carries energy, and is therefore bending spacetime.* It was a heroic effort.

In his famous equations, Einstein specified how to find the universe's gravitational field, given the contents of the universe. Although his best-known equation is $E = mc^2$, physicists use the term "Einstein's equations" to refer to the equations that determine the gravitational field. The equations accomplish this formidable task by showing how to determine the metric of spacetime from a known distribution of matter.[9] The metric you calculate determines the spacetime geometry by telling you how to translate numbers associated with arbitrary scale units into physical distances and shapes that determine the geometry.

With the final formulation of general relativity, physicists could determine the gravitational field and calculate its influence. As with previous formulations of gravity, physicists use these equations to figure out how matter moves in a given gravitational field. For example, they can plug in the mass and position of a big spherical body, such as the Sun or the Earth, and calculate the well-known Newtonian gravitational attraction. In this particular example, the results wouldn't be new—but their meaning would be. Matter and energy bend spacetime, and that bending gives rise to gravity. But general relativity has the further advantage that it incorporates any type of energy—including that of the gravitational field itself—into the distribution of matter and energy. This makes the theory useful even in situations where gravity itself contributes a significant amount of energy.

Because they apply to any distribution of energy, Einstein's equations changed the outlook for cosmologists—historians of the cosmos. Now, if scientists knew the matter and energy content of the universe, they could calculate its evolution. In an empty universe, space would

*Because the gravitational field carries energy, the energy of the field must be taken into account when using Einstein's equations. This makes solving for the gravitational field more subtle than it would be in Newtonian gravity.

117

be completely flat, with no ripples or undulations—no curvature at all. But when energy and matter fill the universe, they distort spacetime, producing interesting possibilities for the universe's structure and behavior over time.

We most definitely do not live in a static universe: as we will soon see, we just might live in a warped, five-dimensional one. Fortunately, general relativity tells us how to calculate their consequences. Just as there are examples of two-dimensional geometries with positive, zero, and negative curvature, there are four-dimensional geometrical configurations of spacetime with positive, zero, and negative curvature, which could arise from appropriate distributions of matter and energy. Later on, when we discuss cosmology and branes in extra dimensions, the distortions of spacetime arising from matter and energy—both in our visible universe and on the branes and in the bulk—will be of critical importance. We'll see that the three types of spacetime curvature (positive, negative, and zero) might be realized in higher dimensions as well.

General relativity has lots of consequences that you can't calculate with Newtonian gravity. Among its many merits, general relativity eliminated the annoying action-at-a-distance of Newtonian gravity, which asserted that an object's gravitational effects would be felt everywhere as soon as it appeared or moved. With general relativity, we know that before gravity can act, spacetime has to deform. This process does not happen instantaneously. It takes time. Gravity waves travel at the speed of light. Gravitational effects can kick in at a given position only after the time it takes for a signal to travel there and distort spacetime. That can never happen more quickly than the time it would take light, which travels as fast as anything we know, to get there. For example, you will never receive a radio signal or a cell phone call sooner than the time it would take for a light beam to travel to you.

Furthermore, physicists were able to use Einstein's equations to explore other types of gravitational field. With general relativity, scientists could describe and study black holes. These fascinating, enigmatic objects form when matter is highly concentrated within a very small volume. In black holes the geometry of spacetime is extremely distorted, so much so that anything entering a black hole

gets trapped inside. Even light cannot escape. Although the German astronomer Karl Schwarzschild discovered that black holes were a consequence of Einstein's equations almost immediately after general relativity's development,* it was not until the 1960s that physicists took seriously the idea that they could be real things in our universe. Today, black holes are well accepted in the astrophysical community. In fact, it looks as though there is a supermassive black hole at the center of every galaxy, including our own. Moreover, if there are hidden dimensions then there exist higher-dimensional black holes which, when big, look like the four-dimensional black holes that astronomers have observed.

Coda

To conclude the story of the GPS system, it turns out that to calculate position to within a meter, we must measure time to better than one part in 10^{13}. The only possible way to get this accuracy is with atomic clocks.

But even if we had perfect clocks, time dilation would slow them down by about one part in 10^{10}. This error, if not corrected, would be a thousand times too big for our desired GPS system. We also have to account for the gravitational blueshift, a general relativity effect associated with a photon traveling in a changing gravitational field, which gives an error at least this great. This and other general relativity deviations would give errors that, if ignored, would build up at a rate greater than 10 km per day.† Ike (and current GPS systems) must correct for these relativistic effects.

Although by now relativity has been well tested and even gives rise to effects that need to be accounted for in practical devices, I do find it fairly remarkable that anyone listened to Einstein at first. He was completely unknown, working in the Bern patent office because he

*He did this on the Russian front while serving with the German army during World War I.

†Neil Ashby, "Relativity and the Global Positioning System," *Physics Today*, May 2002, p. 41.

118

couldn't get a better job. From this unlikely location he proposed a theory that went against the beliefs of all other physicists of his time.

Gerald Holton, a Harvard historian of science, tells me that the German physicist Max Planck was Einstein's first champion. Without Planck, who immediately recognized the brilliance of Einstein's work, it might have taken much longer for it to be recognized and accepted. Following Planck, a few other notable physicists knew enough to listen and pay attention. And shortly afterwards, so did the world.

What to Remember

- The speed of light is constant. It is independent of the speed of an observer.
- *Relativity* modifies our notions of space and time and tells us that we can treat them together as a single *spacetime* fabric.
- Special relativity relates the values of energy, momentum (which tells how an object responds to a force), and mass. For example, $E = mc^2$, where E is energy, m is mass, and c is the speed of light.
- Mass and energy make spacetime curve, and you can think of that curved spacetime as the origin of the gravitational field.

6

Quantum Mechanics: Principled Uncertainty, the Principal Uncertainties, and the Uncertainty Principle

And you may ask yourself,
Am I right? . . . Am I wrong?

Talking Heads

Ike wondered whether Athena was making him watch too many movies or Dieter was talking too much about physics. But whatever the reason, the previous night Ike dreamed he met a quantum detective. Dressed in a fedora, a trench coat, and with a stone-faced expression, the dream detective spoke:

"I knew nothing about her except her name, and that she was standing there before me. But from the moment I set eyes on her I knew Electra would be trouble. When I asked her where she came from, she refused to say. The room had two entrances, and she must have come through one. But Electra whispered hoarsely, 'Mister, forget it. I'll never tell you which.'*

"Although I saw that she was shaking, I tried to pin this lady down. But Electra paced frenetically when I started to approach. She begged me to come no closer. Seeing she was agitated, I kept away. I was no stranger to uncertainty, but this time it had me beat. It looked like uncertainty was going to stick around here for a while."

Quantum mechanics, counterintuitive as it is, fundamentally altered the way scientists view the world. Much of modern science evolved

*The name refers to the electron, not to the character in Greek mythology.

119