Contents

One

The Principle of Relativity

The special theory of relativity was set forth by Einstein in his 1905 paper "On the Electrodynamics of Moving Bodies."[1] The term "special relativity" is used to distinguish the theory from Einstein's theory of gravity, known as general relativity, which he completed ten years later. Except for a glimpse into general relativity in chapter 12, we shall be concerned entirely with special relativity, so from now on I will drop the "special," with the understanding that "relativity" always refers to special relativity.

Einstein based the theory of relativity on two postulates. The first is now known as the principle of relativity. We shall take up the second in chapter 3. Einstein put the principle of relativity this way: "In electromagnetism as well as in mechanics, phenomena have no properties corresponding to the concept of absolute rest." He might have stated it more briefly, and more generally, as "No phenomena have properties corresponding to the concept of absolute rest."

The reason electromagnetism and mechanics get into Einstein's formulation is that the principle of relativity was already a well-known feature of mechanics. It was first enunciated by Galileo, three centuries earlier, and was built into the classical mechanics of Newton. In 1905, however, there was considerable confusion over whether the principle was applicable to electromagnetic phenomena. This accounts for the peculiar title of Einstein's paper, and his emphasis that the principle applied to both mechanical and electromagnetic phenomena. I would guess that he did not explicitly insist that the principle of relativity applied to all phenomena because in 1905 it was still possible to believe that mechanics (and gravity, often viewed at that time as a part of mechanics) and electromagnetism encompassed all the phenomena of nature. Today we know that there are other phenomena (mentioned in chapter 13), but we believe that the principle of relativity applies to all of them.

In this chapter we shall elaborate Einstein's concise statement of the principle of relativity, and then explore how the principle can be used to discover some elementary but not entirely obvious facts about how things behave. A really careful statement of the principle raises some quite subtle

---

[1] A. Einstein, "Zur Elektrodynamik bewegter Körper," *Annalen der Physik* **17** (1905): 891–921.

conceptual issues, which we will note but scrupulously avoid examining in any depth. Such a philosophical study can be entertaining, but it is distracting and of no importance for establishing a working understanding of relativity.

What *is* important is to acquire a sense of how to *use* the principle as a practical tool for enlarging one's understanding of the behavior of moving objects. Using the principle of relativity in such a way may at first be a little unfamiliar, but learning how to do it is quite unrelated to the physical and philosophical subtleties stirred up by an effort to acquire a "deep" understanding of the principle. If one wishes to understand the spectacular and counterintuitive consequences of the straightforward applications of the principle in Einstein's theory of relativity, it is essential to learn first how to apply it to some simpler, less surprising cases.

The principle of relativity is an example of an invariance principle. There are several such principles. They all begin with the phrase "All other things being the same." Then they go on to say:

1. it doesn't matter where you are. (Principle of translational invariance in space)
2. it doesn't matter when you are. (Principle of translational invariance in time)
3. it doesn't matter how you are oriented. (Principle of rotational invariance)

The principle of relativity fits into the same pattern: *All other things being the same*,

4. it doesn't matter how fast you're going if you're moving with fixed speed along a straight line. (Principle of relativity)

"It doesn't matter" means "the rules for the description of natural phenomena are the same." For example the rule describing Newton's force of gravity between two chunks of matter is the same whether they are in this galaxy or another (translational invariance in space). It is also the same today as it was a million years ago (translational invariance in time). The law does not work differently depending on whether one chunk is east or north of the other one (rotational invariance). Nor does the law have to be changed depending on whether you measure the force between the two chunks in a railroad station, or do the same experiment with the two chunks on a uniformly moving train (principle of relativity).

"All other things being the same" raises deep questions. In the case of translational invariance, it means that when you move the experiment to a new place or time you have to move everything relevant; in the case of rotational invariance you have to turn everything relevant. In the case of the principle of relativity, you have to set everything relevant into motion.

If everything relevant turned out to be the entire universe, you might wonder whether there was any content to the principle.

One can thus descend immediately into a deep philosophical abyss from which some never emerge. We shall not do this. We are interested in how such principles work on the practical level, where they are usually unproblematic. You easily can state a small number of relevant things that have to be the same and that is quite enough. When the principle doesn't work, invariably you discover that you have overlooked something else simple that is also relevant. Not only does that fix things up, but often you learn something new about nature that proves useful in many entirely different contexts. If, for example, the stillness of the air was important for the experiment you did in the railroad station, then you had better be sure that when you do the experiment on a uniformly moving train that you do not do it on an open flatcar, where there is a wind, so all other relevant things are not the same. You must do it in an enclosed car with the windows shut. If you hadn't realized that the stillness of the air was important in the station, then the apparent failure of the experiment to work the same way on the open flatcar would teach you that it was.

Invariance principles are useful because they permit us to extend our knowledge to new situations. It is on that quite practical level that we shall be interested in the principle of relativity. It tells us that no experiments that we do can enable us to distinguish between our being in a state of rest or a state of uniform motion. Any set of experiments we perform in a laboratory we choose to regard as being stationary must give exactly the same results as a corresponding set of experiments performed in a laboratory moving uniformly with respect to the first one. The results we get in the new situation, doing experiments in the uniformly moving laboratory, can be inferred from the results we found in the old situation, doing experiments in the stationary laboratory.

It is important both to understand what the principle asserts and to acquire some skill in using it to extend knowledge from one situation to another. But on a deeper level, one can again get bogged down in subtle questions. What do we mean by rest or by uniform motion? We will again take a practical view. Uniform motion means moving with a fixed speed in a fixed direction. More compactly, we say moving with a fixed *velocity*. The term "velocity" embraces both speed and direction of motion. Two boats moving 15 feet per second (f/sec), one going north and the other east, have the same speed but different velocities. I digress to remark that the foot (plural "feet") is a unit of distance (abbreviated "f"), still used in backward nations, equal to 30.48 centimeters. In this book it will be highly convenient to redefine the "foot" to be just a little shorter than the conventional English foot: about 30 centimeters (or, more precisely, 29.9792458 centimeters—98.36 percent of a conventional foot). The reasons for this redefinition will emerge in chapter 3.

It is useful to adopt the convention that a *negative* velocity in a given direction means exactly the same thing as the corresponding positive velocity in the opposite direction: −10 f/sec east is exactly the same as 10 f/sec west. Note also that in the definition of uniform motion, a fixed direction is just as important as a fixed speed: something moving with fixed speed on a circular path is not moving uniformly.

A state of nonuniform motion can easily be distinguished from a state of rest or uniform motion. You can clearly tell the difference between being in a plane moving at uniform velocity and being in a plane moving in turbulent air; between being in a car moving at uniform velocity and in one that is accelerating or cutting a sharp curve or on a bumpy road or screeching to a halt. But you cannot tell the difference (without looking out the window) between being on a plane flying smoothly through the air at 600 f/sec and being on a plane that is stationary on the ground.

In working with the principle of relativity, one uses the term *frame of reference*. A frame of reference (often simply called a "frame") is the system in terms of which you have chosen to describe things. For example, a flight attendant walks toward the front of the airplane at 3 f/sec in the frame of reference of the airplane. You start at the rear of the plane and want to catch up with him so you walk at 6 f/sec in the frame of the plane. If the plane is going at 700 f/sec, then in the frame of reference of the ground this would be described by saying that the cabin attendant was moving forward at 703 f/sec, and you caught up by increasing your speed from 700 to 706 f/sec. One of the many remarkable things about relativity is how much one can learn from considerations of this apparently banal variety.

Another important term is *inertial frame of reference*. "Inertial" means stationary or uniformly moving. A rotating frame of reference is not inertial, nor is one that oscillates back and forth. We will almost always be interested only in inertial frames of reference and will omit the term inertial except when we wish to contrast uniformly moving frames of reference to frames that move nonuniformly.

How do you know that a frame of reference is inertial? This is just another way of posing the deep question of how you know motion is uniform. It would appear that you have to be given at least one inertial frame of reference to begin with, since otherwise you can ask "Moving uniformly with respect to what?" Thus if we know that the frame in which a railroad station stands still is an inertial frame, then the frame of any train moving uniformly through the station is also an inertial frame. But how do we know that the frame of reference of the station is inertial?

Fortunately, there is a simple physical test for whether a frame is inertial. In an inertial frame, stationary objects on which no forces act remain stationary. It is this failure of a stationary object (you) to remain stationary (you are thrown about in your seat) that lets you know when the plane or car you are riding in (and the frame of reference it defines) is moving uniformly and when it is not. In our cheerfully pragmatic spirit, we will set aside the deep question of how you can know that no forces act. We will be content to stick with our intuitive sense of when the motion of an airplane (train, car) is or is not capable of making us seasick.

When specifying a frame of reference you can sometimes fall into the following trap: suppose you have a ball that (in the frame of reference you are using) is stationary before 12 noon, moves to the right at 3 f/sec between 12 p.m. and 1 p.m., and to the left at 4 f/sec after 1 p.m. By "the frame of reference of X" (also called the *proper frame* of X), one means the frame in which X is stationary. Now there is no *inertial* frame of reference in which the ball is stationary throughout its whole history. If you want to identify an inertial frame of reference as "the frame of reference of the ball," you must be sure to specify whether you mean the inertial frame in which the ball was stationary before 12, or between 12 and 1, or after 1. Depending on the time, three different inertial frames can serve as the frame of reference of the ball. Similarly for the Cannonball Express, which defines one inertial frame of reference as it zooms along a straight track at 150 f/sec from New York to Chicago, and quite another as it zooms along the same track at the same speed on the way back. The frame of reference of an airplane buffeted by high winds may never be inertial. Nor is the frame of reference of the Cannonball as it moves with fixed speed along a curved stretch of track.

Here is another, more subtle trap that many people (including, I suspect, some physicists) fall into: people sometimes take the principle of relativity to mean, loosely speaking, that the behavior of a uniformly moving object should not depend on how fast it is moving, or, to put it slightly differently, that motion with uniform velocity cannot affect any properties of an object. This is simply wrong. The principle of relativity only requires that if an object has certain properties in a frame of reference in which the object is stationary, then if the same object moves uniformly, it will have the same properties *in a frame of reference that moves uniformly with it.* But the properties that an object can have in a fixed frame of reference can certainly depend on the speed with which it moves uniformly in that frame. To take a silly example, when the object moves it has a nonzero speed, but when it is stationary its speed is zero. You could, of course, object that speed is not a property inherent in an object, but specifies a relation between the object and the frame of reference in which it has that speed. This is fine. But the nature of the trap is then that many properties that might appear to be inherent in an object turn out, on closer examination, to be relational. We shall see many examples of this.

A less trivial example is provided by the *Doppler effect.* If a yellow light moves away from you at an enormous speed, the color you see changes from yellow to red; if it moves toward you at an enormous speed, the color changes from yellow to blue. So the color of an object in a fixed frame of reference can depend on whether it is moving or at rest, and in what direction it is moving. What the principle of relativity does guarantee is that if a light is seen to be yellow when it is stationary, then when it moves with uniform velocity it will still be seen as yellow *by somebody who moves with that same velocity.*

We shall be almost exclusively interested in some simple practical applications of the principle of relativity. To apply the principle, it is essential to acquire the ability to visualize how events look when viewed from different inertial frames of reference. A useful mental device for doing this is to examine how a single set of events would be described by various people moving past them, in trains moving uniformly with different speeds.

We will be applying the principle of relativity to learn some quite extraordinary things by examining the same sets of events in different frames of reference. Some of the things we shall learn in this way are so surprising that they are hard to believe at first. You are more likely to conclude that you must have made a mistake in applying the principle. So it is quite essential to begin by acquiring some skill in using the principle of relativity to learn some things that you might not have known before, which, though not obvious, are also not astonishing. The general procedure for doing this is always the same: *Take a situation which you don't fully understand. Find a new frame of reference in which you do understand it. Examine it in that new frame of reference. Then translate your understanding in the new frame back into the language of the old one.*

Here is a very simple example. Newton's first law of motion states that in the absence of an external force a uniformly moving body continues to move uniformly. This law follows from the principle of relativity and a very much simpler law. The simpler law merely states that in the absence of an external force a stationary body continues to remain stationary.

To see how the more general law is a consequence of the simpler one, suppose we only know the simpler law. The principle of relativity tells us that it must be true in all inertial frames of reference. If we want to learn about the subsequent behavior of a ball initially moving at 50 f/sec in the absence of an external force, all we have to do is find an inertial frame of reference in which we can apply the simpler law. The frame we need is clearly the one that moves at 50 f/sec in the same direction as the ball, since in that frame of reference the ball is stationary. To put it more concretely, think of how the ball looks from a train moving at 50 f/sec alongside it. In the frame of reference of the train, the ball is stationary and we can apply the law that in the absence of an external force a stationary body remains stationary. But anything that is stationary in the train frame moves at 50 f/sec in the frame of reference in which we originally posed the problem. We conclude that since the ball remains stationary in the train frame in the absence of an external force, in the original frame it must continue to move at 50 f/sec in the absence of an external force.

So starting with the fact that undisturbed stationary objects remain stationary, we have used the principle of relativity to establish the much more general fact that undisturbed uniformly moving objects continue to move with their original velocity. (At the risk of complicating something simple, I feel obliged to remark that in reaching this conclusion we have implicitly assumed that if an object is undisturbed in one inertial frame of reference, then it is undisturbed in any other inertial frame of reference—that the condition of no force acting on an object is an *invariant* condition independent of the frame of reference in which the object is described. Since such forces can be associated with jet engines being on or off, springs being compressed or slack, etc., this is a reasonable assumption.)

If you already knew Newton's first law of motion, you might not be impressed at the power of this line of thought, so let's examine a case where what we learn might not be quite so familiar. Suppose we have two identical perfectly elastic balls. Identical elastic balls have the property that if you shoot them directly at each other with the same speed, then after they collide each bounces back in the direction it came from with the same speed that it had before the collision. Question: What happens if one of the balls is at rest and you shoot the other one directly at it?

There is a long tradition of answering such questions by invoking the conservation of energy and momentum. If you happen to know how to use such conservation laws, you should forget this for now—we shall return to the use of conservation laws in collisions in chapter 11. At this stage it is both entertaining and instructive to understand how this and many related questions can be answered using nothing but the principle of relativity. In learning how to use the principle in this way you will acquire a conceptual skill that will be essential in understanding everything that is to follow. Indeed, answering such questions using the principle of relativity provides a deeper insight than answering them by applying conservation laws.

To figure out what happens, using only the principle of relativity, first draw a picture illustrating the rule you know: when the balls move at each other with equal speeds, they simply rebound with the same speeds. This is shown in the upper part of figure 1.1. Then draw a picture of the new situation, shown in the lower part of figure 1.1. For concreteness I've taken the original speed of the moving ball to be 10 f/sec. We want to know what goes in the box in figure 1.1 with the question mark in it.

To understand what happens in the unknown case, consider it to be taking place along the tracks in a railroad station. The white ball moves to

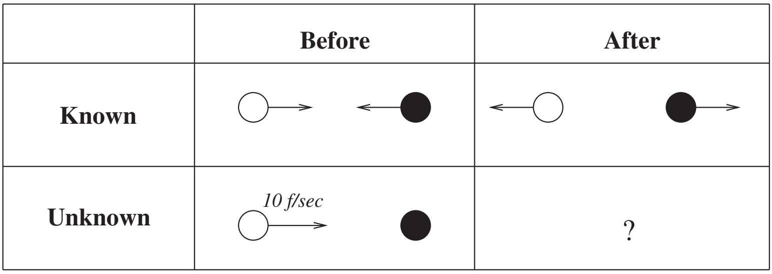

Figure 1.1

the right along the tracks at 10 f/sec toward the stationary black ball. This is shown again in the upper part of figure 1.2. Now think about how this would look if we were describing it from the frame of reference of the train moving through the station to the right at 5 f/sec, shown in the middle part of figure 1.2. Since the white ball covers 10 feet of track per second, and the train covers 5 feet of track per second, every second the white ball gains 5 feet on the train. So in the frame of reference of a train moving to the right at 5 f/sec, the white ball is moving to the right at 5 f/sec. Since the black ball is stationary with respect to the tracks, in the train frame it moves to the *left* at 5 f/sec, just as the tracks do. Therefore in the frame of this particular train the unknown situation before the collision, pictured just above the picture of the train in figure 1.2, becomes an instance of the known situation, pictured just below the picture of the train, in which the balls approach each other with the same speed. But the principle of relativity assures us that any experiment we do with the two elastic balls must have the same outcome in any inertial frame of reference. Since the two balls are moving at each other with the same speed—5 f/sec—in the train frame, after the collision they must bounce away from each other, each still moving at 5 f/sec in the train frame. This is also pictured just below the picture of the train.

Now all that remains is to translate this train-frame answer back to the original frame of reference—the station frame. After the collision the white ball moves to the left at 5 f/sec in the train frame, so it must be stationary in the station frame. After the collision the black ball moves to the right at 5 f/sec in the train frame, so it must be moving to the right at 10 f/sec in the station frame. This is pictured in the lowest part of figure 1.2.

So we have used the principle of relativity to learn something new about identical elastic balls: if one is at rest and the other bumps it head-on, then the moving one comes to a complete stop and the stationary one moves off with the velocity the formerly moving one originally had. This is a

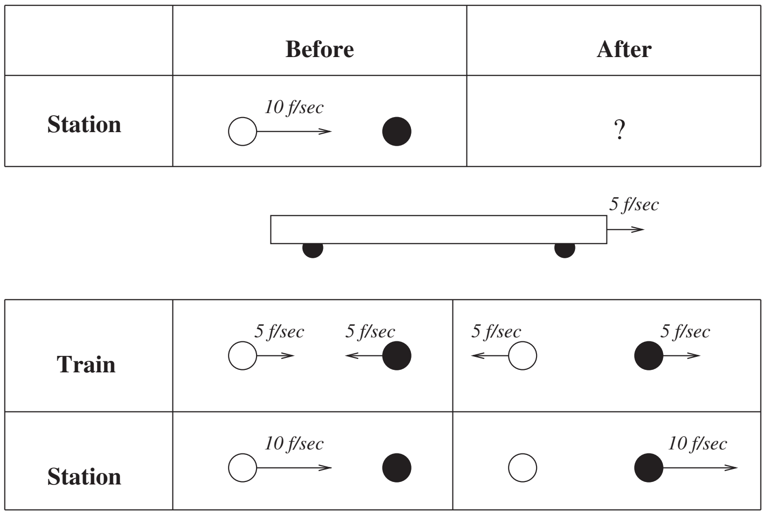

Figure 1.2

fact familiar to all players of billiards, but not many of them realize that it is simply a consequence of the much more obvious fact (less frequently encountered in billiards) that when two balls collide head-on with equal and opposite speeds, each bounces back the way it came with its original speed.

This is a dramatic illustration of the power of the principle of relativity. There is nothing very interesting about watching two balls approaching each other with equal speeds and then bouncing back, each with the same speed. If the balls are identical and sufficiently elastic, what else would you expect them to do? On the other hand, when you see a moving ball hit a stationary ball and immediately come to a complete stop, while the formerly stationary ball goes zooming off at the same speed as the originally moving ball, this can have a spectacular feel to it. One can't help wondering how the moving ball manages to come to a perfect stop and how the stationary ball manages to acquire *exactly* the same speed as the one that hit it. This mystery is solved, when you realize that the spectacular collision is just the boring one, viewed from an appropriately moving frame of reference. If you can train your imagination to experience this connection at some visceral level, then you will have mastered the principle of relativity.

Another example is shown in figure 1.3. Two identical sticky balls have the property that if they are fired directly at each other with equal speeds, then they stick together upon collision and the resulting compound ball is

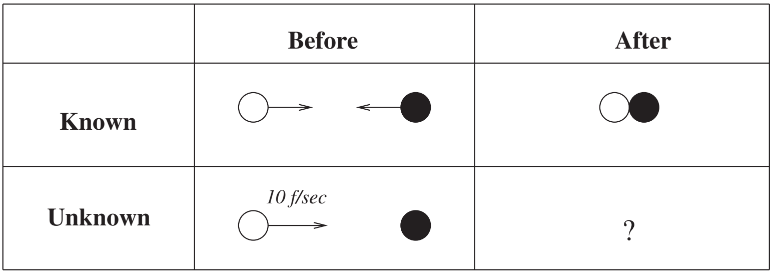

Figure 1.3

stationary. If a sticky ball is fired at 10 f/sec directly at another identical sticky ball that is stationary and the two stick together, with what speed and in what direction will the compound ball move after the collision?

We can again answer the question using only the principle of relativity, by viewing the initially moving white ball and initially stationary black ball from the frame of reference of a train in which both are moving with the same speed but in opposite directions, as shown in figure 1.4. As in our earlier example, such a train moves along the direction of motion of the white ball but only at 5 f/sec. In the train frame the situation before the collision is the one we know about: the balls move toward each other at the same speeds. Therefore in the train frame we know that after the collision the compound ball is stationary. But since the train moves down the tracks at 5 f/sec and the compound ball is stationary in the train frame, in the track frame it will move down the tracks at 5 f/sec—the same speed as the train moves in the track frame. This solves the problem: when the moving ball strikes the stationary ball, the resulting compound ball moves at half the original speed of the moving ball.

A third example is given in figure 1.5. This one has the virtue that it will not be obvious how to solve it if you happen to know about conservation of momentum, but it is easily solved using the principle of relativity. Suppose we have two elastic balls, but one of them is very big and the other is very small. If the big ball is stationary and the small ball is fired directly at it, the small ball simply bounces back in the direction it came from with the same speed, and the big ball stays at rest. (Think of throwing a table-tennis ball directly at a bowling ball.) With what speed will each ball move after the collision, if the small ball is stationary and the big ball is fired directly at it with a speed of 10 f/sec?

We wish to examine the initial situation in a frame of reference in which the big ball is stationary, so we must now view the collision in the frame of a train moving, with the big ball, at 10 f/sec to the left, as shown in figure 1.6. In that frame the small ball will move at 10 f/sec to the right,

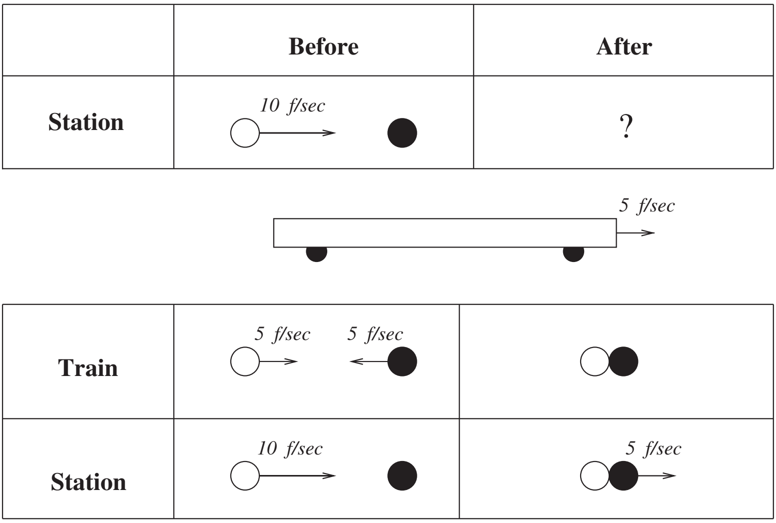

Figure 1.4

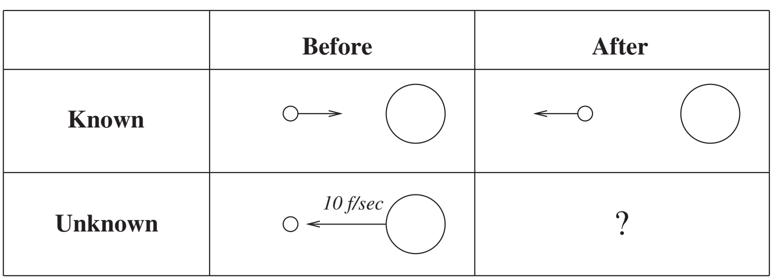

Figure 1.5

and the situation before the collision is the one we understand. So in the train frame we know that after the collision the big ball will remain stationary and the small ball will move at 10 f/sec to the left. Returning to the description in the station frame, we note that after the collision the big ball moves with the train, at 10 f/sec to the left. The little ball, however, moves at 20 f/sec to the left, since in each second it gains 10 feet on the train, which has itself moved 10 feet to the left. So if the little ball is initially stationary, then after a collision with the big ball it moves off at *twice* the speed of the big one.

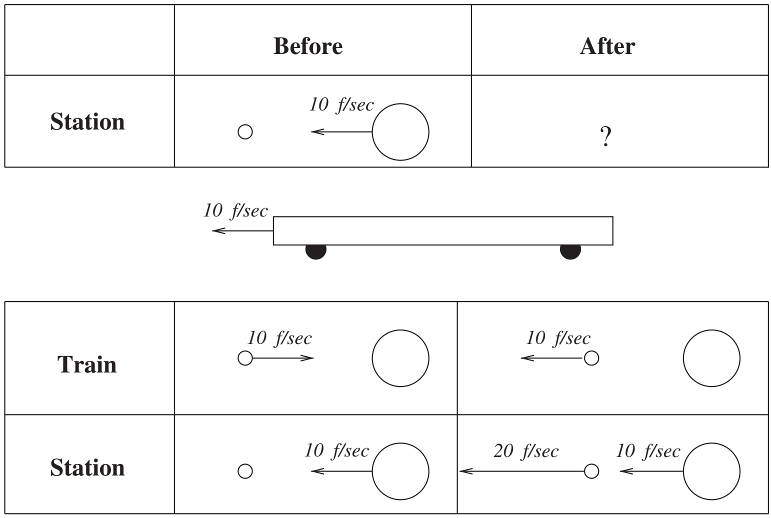

Figure 1.6

There is another interesting case to examine here, shown in figure 1.7. What happens if the big and little ball approach each other with the *same* speed—say 5 f/sec. In that case the train providing the frame of reference in which we know the answer moves, with the big ball, to the left at 5 f/sec, so the little ball moves to the right at 10 f/sec in the train frame. After the collision the big ball remains stationary in the train frame, while the little ball moves to the left at 10 f/sec. So back in the station frame the little ball moves to the left at 15 f/sec, with *triple* its original speed.

I note in passing that you can see a spectacular demonstration of this by placing the little ball, for example a tennis ball, at the very top of the big ball, for example a basketball, and then dropping them on a hard surface very carefully, so that the little ball does not roll off the top of the big one. When the big ball hits the floor it reverses its direction of motion without a change in speed, so for a very brief moment the big ball is moving up and the little ball is moving down, both going at the same speed. Immediately after that, the little ball flies up at nearly three times its original speed. As it happens, the height reached by a ball moving up is proportional to the *square* of its initial speed, so if losses due to various kinds of friction are unimportant, the little ball can shoot up to almost nine times the height from which it was originally dropped!

Seeing this has the wonderful feel of watching a good magic trick. And like a really good magic trick, when you figure out why it works—

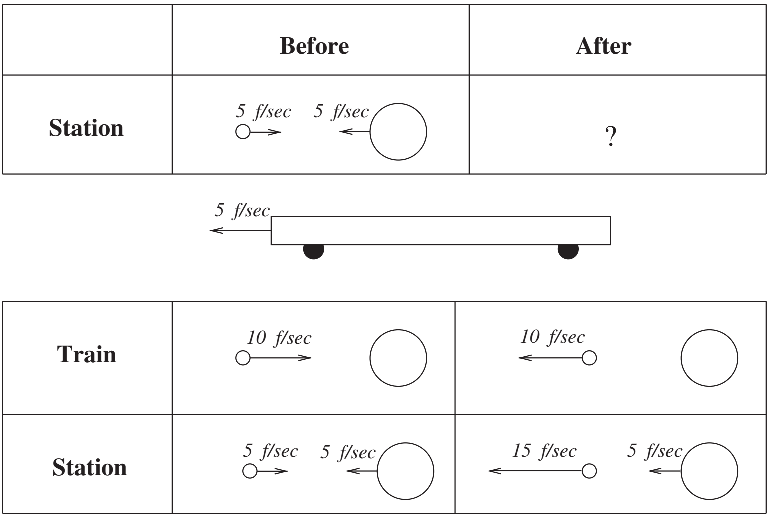

Figure 1.7

figure 1.7 gives the explanation—your appreciation for the trick is only enhanced by its simplicity.

I hope these examples will give you a feeling for how the principle of relativity is actually used, and for the power it can have to predict behavior under apparently unfamiliar conditions. Before starting to apply it under genuinely unfamiliar conditions, we must look a little more closely at some of the reasoning we used in these simpler examples.

INDEX