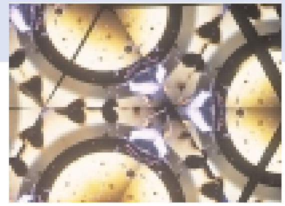
CHAPTER 19



NEW GEOMETRIES FOR A NEW UNIVERSE

As an organized body of knowledge, geometry consists of statements (*theorems* and *corollaries*) logically derived from other statements (*postulates,* more commonly called *axioms*) that are assumed to be true. About 500 B.C., Euclid presented five postulates from which he developed a large body of theorems, a system that we call **Euclidean geometry.**

From then until late in the nineteenth century, Euclidean geometry—including its extension to three dimensions—was thought to be the only mathematics of space. Its theorems were thought to be truths about the world in which we live. No one could imagine a different geometry.

New kinds of *non-Euclidean* geometry were conceived during the nineteenth century and came into their own only in the twentieth century, when one of them became the basis for a major revolution in physics and cosmology—the theory of relativity.

A **non-Euclidean geometry** is any collection of postulates, theorems, and corollaries for geometry (concerning points, lines, circles, and angles) that differs from the collection formulated by Euclid.

EUCLIDEAN GEOMETRY

Euclid's five postulates concern points, lines, circles, and angles, and presuppose notions about what it means for a point to be on a line and for two lines to meet in a point. The postulates were intended to be absolute, self-evident truths. However, they are not postulates about points as pencil dots or lines drawn on paper with a ruler, but instead attempt to characterize the ideal concepts behind these physical realizations. Although the points and lines of our (and Euclid's) experience are the inspiration for the postulates, and Euclid gave fuzzy definitions of them (e.g., a point is "that which has no dimension"), the "points" and "lines" of the postulates are basic undefined terms. Their properties are described implicitly by the postulates since only the postulates are used in reasoning about the concepts. Any objects that satisfy the postulates can be taken to be "points" and "lines."

Paraphrased somewhat, Euclid's postulates are as follows:

1. Two points determine a line.
2. A line segment can always be extended.
3. A circle can be drawn with any center and any radius.
4. All right angles are equal.
5. If l is any line and P any point not on l, then there exists exactly one line m through P that does not meet l.

These five statements were supposed to be absolute, self-evident truths. The first four are simple statements, sufficiently unrelated that it can be shown fairly easily that each is *logically independent* of the others. What this means is that we cannot conclude from the other three postulates whether the remaining postulate is false (in which case our system would be self-contradictory) or true (in which case the remaining postulate is superfluous).

A statement is **logically independent** of a collection of other statements if neither it nor its negation can be deduced from them.

The fifth postulate concerns what are known as *parallel lines:*

Two lines are **parallel** if they do not meet.

You may remember from high school geometry that in Euclidean geometry in the plane, parallel lines are an equal distance apart at all points. This fact of Euclidean geometry can be proved with the use of additional postulates about distance. However, the definition of "parallel" is that lines do not intersect; parallel lines in non-Euclidean geometries are *not* equidistant everywhere.

The fifth postulate is usually referred to as *Euclid's parallel postulate.*

Euclid's parallel postulate states that for any line l and any point P not on l, there is exactly one line m through P that is parallel to (that is, does not meet) l.

Many early geometers thought that this postulate is not independent of the first four but instead is a logical consequence of them. In fact, Euclid himself may have thought his parallel postulate to be an unnecessary assumption for his geometry, for he derived nearly 30 theorems before using it.

The long history of attempts to derive Euclid's parallel postulate as a consequence of the first four postulates is a fascinating story of failures. Whenever someone proposed a proof, it was found to be tacitly based on some extra assumption in addition to the first four postulates—an assumption *logically equivalent* to the parallel postulate. The reasoning was circular, hence invalid.

Two statements are **logically equivalent** if each can be deduced from the other.

Some hidden assumptions logically equivalent to the parallel postulate are:

- The sum of the angles of a triangle equals 180°.
- There is exactly one circle through any three points that are not on the same line.
- Parallel lines are equidistant.

ELLIPTIC GEOMETRY

G. F. Bernhard Riemann (1826–1866) (see Figure 19.1), a young German mathematician, analyzed postulate 2 ("A line segment can always be extended"). He observed that such a property of lines should be distinguished from "A line is infinite." That is, *unboundedness does not imply infinite extent.*

What Riemann had in mind was the geometry of the surface of the earth, the **spherical geometry** used in navigation. The "points" of this geometry are the points on the surface of the sphere. But what are the "lines"?

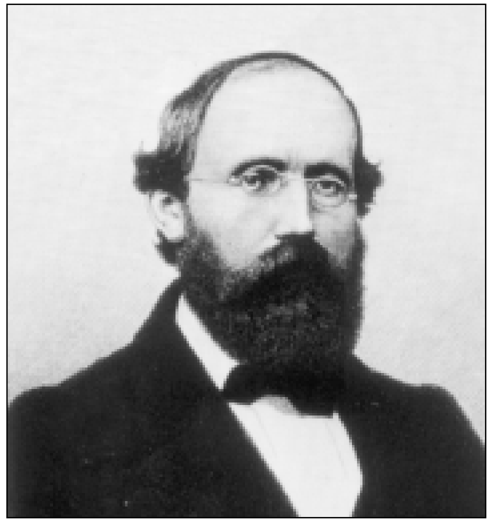


FIGURE 19.1 G. F. Bernhard Riemann

Think of the surface of the earth. Going "straight" along what you would imagine was a line, you can travel another mile and another mile, and so on, and you would eventually return to your starting point. You have traveled on a finite path, but one that is unbounded—that is, you can keep traveling on and on. When Riemann investigated the consequences of a line coming back on itself, he came to the conclusion that interpreting postulate 2 as referring to *unbounded* lines rather than infinite ones opens the door to a geometry that abandons Euclid's parallel postulate in a striking way. Riemann replaced Euclid's parallel postulate with:

Postulate E: Every two lines intersect.

Postulate E leads to *elliptic geometry.*

An **elliptic geometry** is one in which there are no parallel lines.

From Euclidean geometry in the plane, you are accustomed to the line through two points as determining "the shortest distance between two points." That useful property suggests that a good interpretation of "line" for the surface of the sphere might be the analogous shortest route on the surface of the sphere.

However, there are no "straight" lines that lie on the surface of the sphere. The shortest distance between two points on the surface of the sphere would be along a tunnel between them, but such a route would be "out of bounds": we must stay on the *surface* of the sphere and consider "lines" made up of points on that surface.

A curve (possibly straight) that gives the shortest path between two points on a surface is called a **geodesic.**

If we intersect the sphere in Figure 19.2 with a plane through *A* and *B*, the cross section is a circle passing through the given points; and the shorter of the two arcs from *A* to *B* of this circle is a candidate for the shortest path from *A* to *B*. Every plane section of the sphere is a circle, each with a different curvature. The larger the circle, the less curving; the less curving, the shorter the path between *A* and *B*. Thus, the largest circle obtained as a cross section of the sphere gives rise to the shortest path. The largest circle is the *great circle,* obtained by the plane determined by *A, B,* and the center of the sphere (see Figure 19.2).

A **great circle** on a sphere is a circle whose plane includes the center of the sphere.

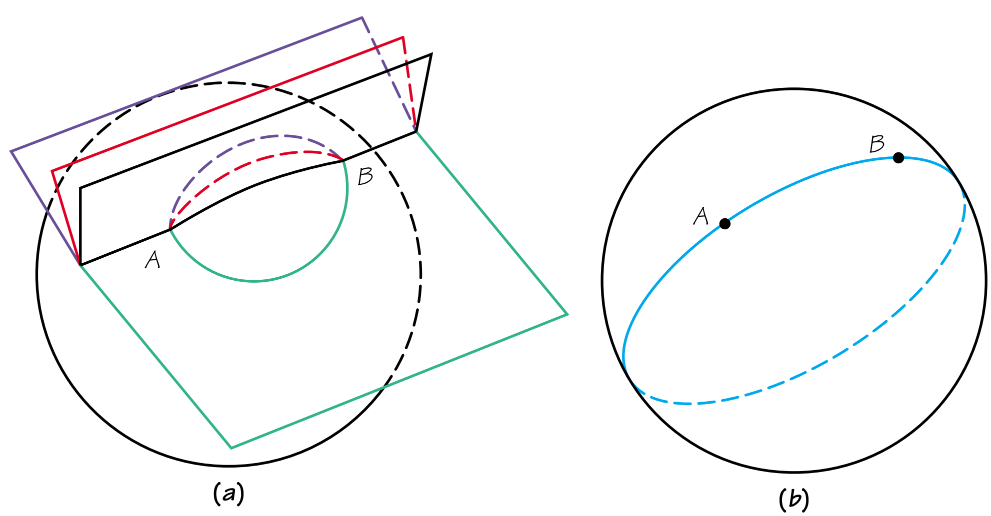


FIGURE 19.2 (a) The planes through *A* and *B* intersect the sphere in circles, the largest of which has the smallest curvature and the smallest distance from *A* to *B*. (b) The great circle through *A* and *B*.

Great circles are familiar to pilots who fly long distances and ship captains on long voyages. If an airplane is on the equator and the pilot wishes to fly to another point on the equator, the pilot would simply fly along the equator, which is a great circle. The great circle "lines" of the sphere can indeed be extended indefinitely, since you can keep on going round and round the earth along one. However, if the airplane is at 10° north latitude and the pilot wishes to fly to a destination at the same latitude, then the shortest distance requires going farther north. An airplane flying from New York to Naples, both at the same latitude, would travel quite far north in the Atlantic Ocean, while the shortest flight path from Washington, D.C., to Ho Chi Minh City, Vietnam, is almost directly over the North Pole. In the polar view of Figure 19.3, the shortest path appears as almost a straight line segment, as the great circle between the two cities is almost edge-on to you.

Great circles look curved when we observe the earth from space, but they do not look curved as we move on the surface of the earth. A ship moving directly east along the equator is following a curved path around the earth; but as far as the crew is concerned, the ship is always headed "straight" east.

Riemann's predecessors certainly realized the practicalities of navigation on the sphere, and about great circles intersecting, but they did not think of the geometry of the surface of the sphere as an alternative to Euclidean geometry—despite the fact that we live not on a plane but on a sphere.

In the geometry of the surface of the sphere, if we take the "lines" to be the great circles, then there are no parallel "lines." Every pair of "lines" (great circles) intersect, so postulate E holds.

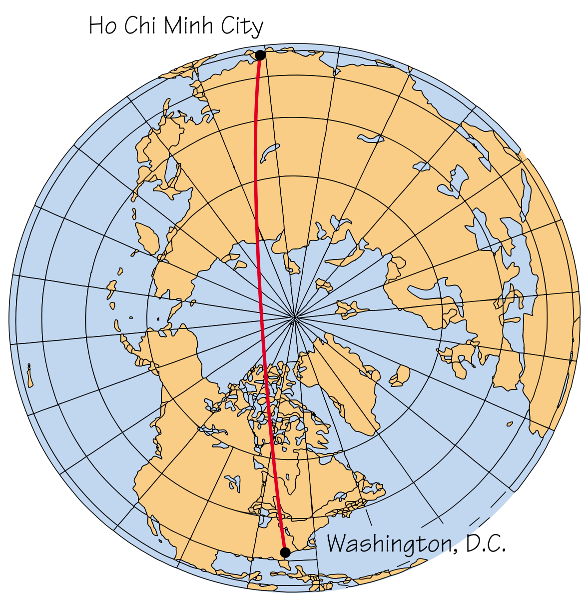


FIGURE 19.3 The shortest path between two cities is an arc of a great circle.

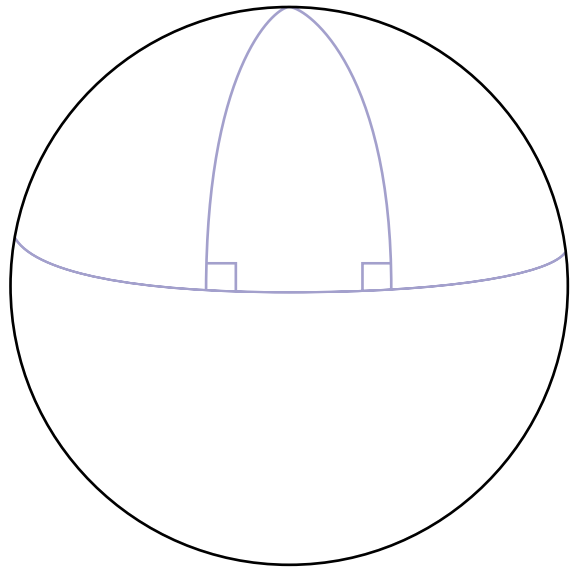


FIGURE 19.4 In elliptic geometry, a triangle can have two or more right angles.

The geometry of the surface of a sphere has other notable differences from Euclidean geometry. Postulate 1 fails for points that are directly opposite each other on the sphere, since not one but an infinite number of elliptic lines (great circles) go through two such points. On the surface of a sphere, triangles can have two or even three right angles: just put two vertices on the equator and one at a pole (see Figure 19.4). In fact, a general theorem of elliptic geometry states that *in an elliptic geometry, the sum of the angles of any triangle is greater than 180°.*

HYPERBOLIC GEOMETRY

The first non-Euclidean geometry was formulated not by Riemann but by Nikolai Ivanovich Lobachevsky (1792–1856) and János Bolyai (1802–1860). (Carl Friedrich Gauss [1777–1855] investigated hyperbolic geometry before Lobachevsky and Bolyai, but did not publish his results.) Unlike many of their predecessors, they did not try to derive Euclid's parallel postulate from the other postulates. Instead, working alone, each decided that Euclid's parallel postulate must be logically independent of the other postulates. They realized that this means that those postulates really have nothing to say about the existence or nonexistence of parallels. Thus, including Euclid's parallel postulate with the other postulates cannot lead to any inconsistencies with them. Similarly, if they were to leave out Euclid's parallel postulate and include instead an alternative postulate about parallels—even one contradictory to Euclid's parallel postulate—they should once again get a consistent system of postulates. They both chose the same alternative postulate and proceeded to invent and explore the resulting "non-Euclidean" geometry, proving theorems from its postulates in the same style that Euclid had proved theorems from his (see Spotlight 19.1).

The alternative postulate that Lobachevsky and Bolyai chose was:

Postulate H: If l is any line and P is any point not on the line, then there exists more than one line through P not meeting l.

The use of postulate H leads to an entirely new system of theorems and corollaries, which we now call *hyperbolic geometry.*

A **hyperbolic geometry** is one in which for any given line and a point not on the line, more than one line passes through the point and is parallel to the given line.

Some of the theorems in hyperbolic geometry are exactly the same as in Euclidean geometry, because theorems derived only from postulates 1 through 4 must be valid in both systems. However, hyperbolic geometry provides some new and very surprising theorems.

Lines in hyperbolic geometry behave differently from what our intuition about Euclidean lines leads us to expect. In representing hyperbolic geometry in the plane of the page, it is useful to represent hyperbolic lines as curved. However, just as great circles appear straight to people living on a sphere, hyperbolic lines would appear perfectly straight, and to be geodesics (giving shortest distances), to people living on a hyperbolic surface.

Figure 19.5 shows a line l (shown curved) and a point P not on l. We drop a perpendicular from P to l, calling A the foot of the perpendicular. Now, consider the line PE, which is perpendicular to PA. The line PE is parallel to l (it does not intersect it). Under Euclid's parallel postulate, it would have to be the only parallel to l through P. However, if we assume H instead, then there is *another* "line" m that passes through P and is parallel to l. It must make a smaller angle with PA, as shown on the left in Figure 19.5 (if it too made a right angle, then it couldn't be a different line from PE, by postulate 4).

A key observation in Figure 19.5 is that with m making an acute angle with PA on the left side of PA, and necessarily an obtuse angle on the right side, there must be yet another line *n* making the same acute angle with *PA* on the right side. By symmetry, *n* too must be parallel to *l*.

From this argument comes the first astonishing conclusion, that through *P* there are *infinitely many* parallels to line *l*. This is clear as soon as you consider all the lines through *P* and divide them into two classes. One class contains *PA* and lines making a smaller angle with *PA* than *n* and *m* do. The other class contains *PE*, *n*, *m*, and all lines lying between *PE* and *n* and between *PE* and *m* (they lie in the shaded region in Figure 19.5). All the lines in the second class are parallel to *l*.

Using similar reasoning, Bolyai and Lobachevsky discovered many unusual theorems, of which we list three that were important for their shock value to

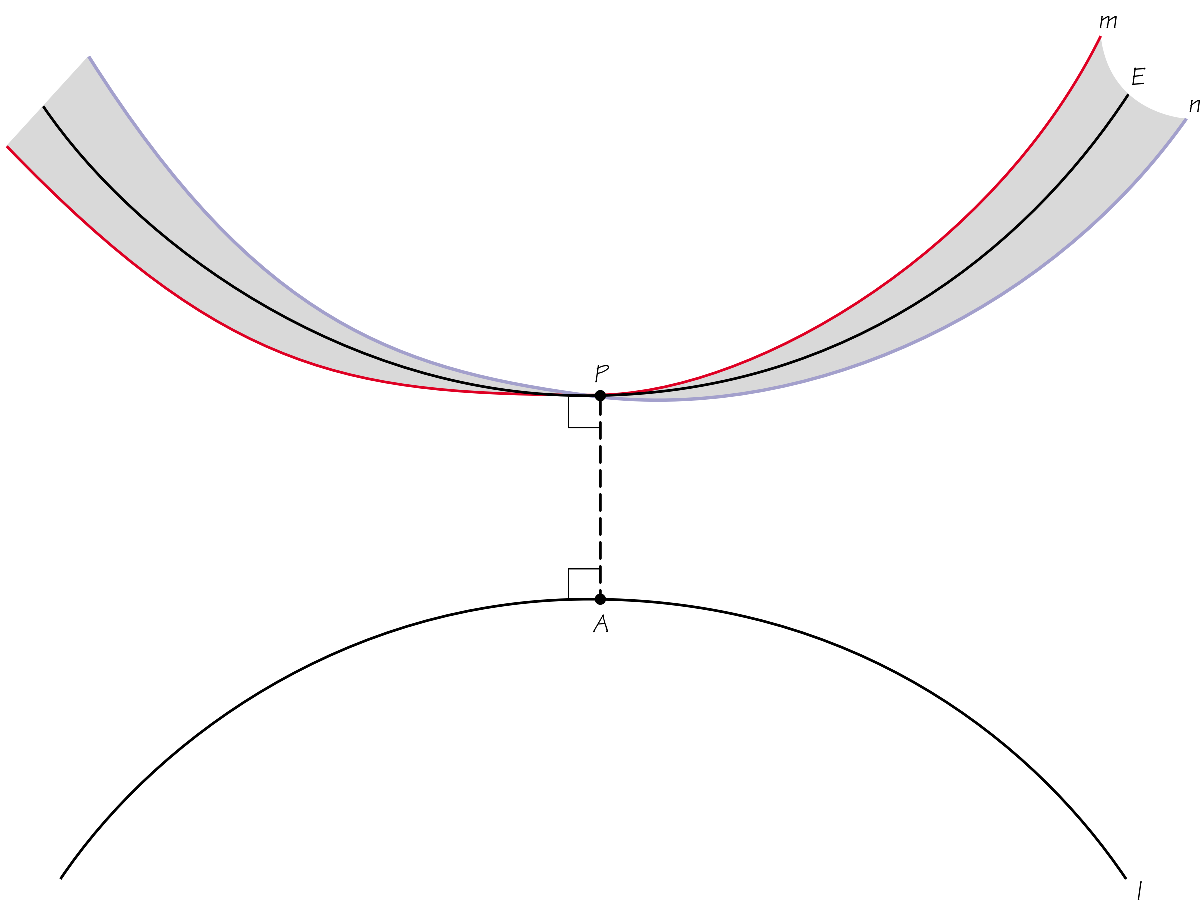


FIGURE 19.5 In hyperbolic geometry there is more than one parallel through a point *P* not on a given line *l*.

the dubious mathematicians who listened to their theories (Figure 19.6 illustrates theorems 1 and 3):

1. The sum of the angles in any triangle is less than 180°.
2. Similar triangles are congruent—that is, triangles having the same shape (angles) must be the same size; the size of a triangle depends on the sum of its angles.
3. Given two parallel lines, there exists a third line perpendicular to one and parallel to the second.

Hence, in each of the three geometries we have investigated, we have different sums for the angles in a triangle:

- $<180°$ in hyperbolic geometry (angle sums range between 0° and 180°)
- $=180°$ for all triangles in Euclidean geometry (sometimes called *parabolic geometry*) (see Spotlight 19.2)
- $>180°$ in elliptic geometry (angle sums range between 180° and 540°)

Thus, having a constant angle sum for all triangles dramatically distinguishes Euclidean geometry from elliptic geometry and hyperbolic geometry. In both elliptic geometry and hyperbolic geometry, any two triangles with the same area have the same angle sum.

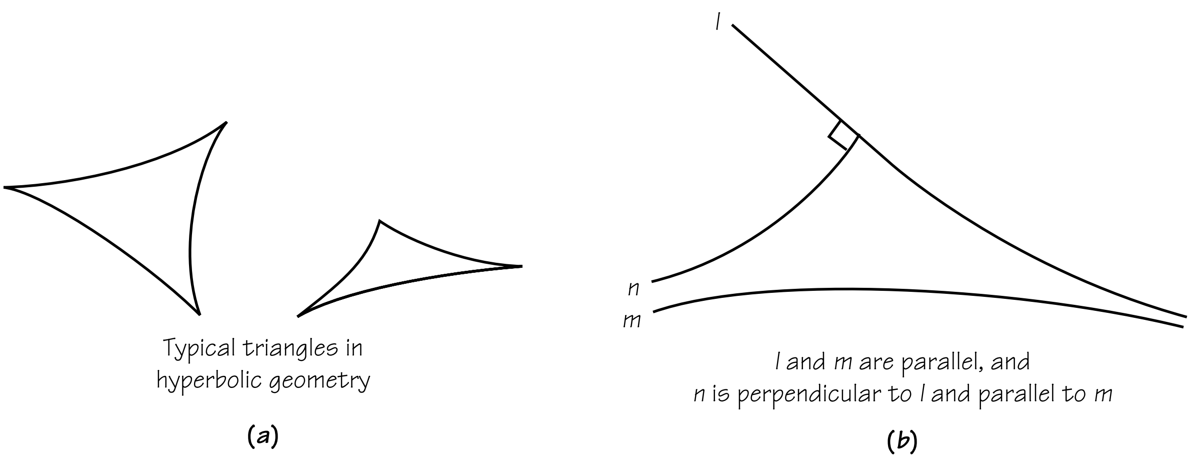


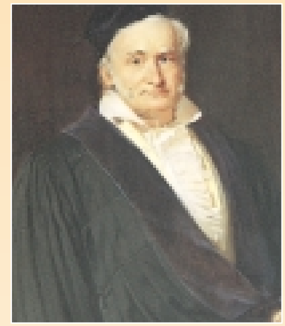
FIGURE 19.6 (a) Typical triangles in hyperbolic geometry; the sum of the angles of any triangle is less than 180°. (b) Given two parallel lines *l* and *m*, there exists a third line *n* perpendicular to *l* and parallel to *m*.

Spotlight 19.1

The Discovery of Non-Euclidean Geometry

Nikolai Ivanovich Lobachevsky

János Bolyai



Carl Friedrich Gauss

Nikolai Ivanovich Lobachevsky (1792–1856) and János Bolyai (1802–1860) independently discovered non-Euclidean geometry. Lobachevsky was the first to publish an account of it (1829), which he first called "imaginary geometry" and later "pangeometry." His work attracted little attention, largely because it was written in Russian and the Russians who read it were very critical.

Bolyai published his work as a 26-page appendix to a book (the *Tentamen,* 1831) by his mathematician father Wolfgang, who proudly sent the work by his son to Carl Friedrich Gauss (1777–1855), the leading mathematician of his day. Gauss replied to Wolfgang that he himself had earlier discovered non-Euclidean geometry! From correspondence and private papers that became available after his death, we know that Gauss's claim was correct, though Bolyai and Lobachevsky deserve credit for having the courage to publish their discoveries.

MODELS FOR HYPERBOLIC GEOMETRY

For Euclidean geometry, we have as a natural *model* the plane, with "points" and "lines" interpreted as the points and lines that you are accustomed to.

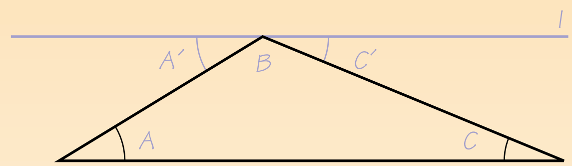
> An example that satisfies a collection of axioms is called a **model** of the axioms.

Spotlight 19.2 Angle Sums in a Triangle

The following simple proof shows that the angle sum in a triangle is equal to a straight angle. Let triangle *ABC* be an arbitrary triangle. Our goal is to prove that $A + B + C =$ a straight angle ($=180°$).

The Proof

At vertex *B*, draw a line *l* parallel to side *AC*. Then, by one of the first properties of parallels, which states that when parallel lines are cut by a transversal, the "alternate interior angles" are equal, angle $A =$ angle A'. Likewise, angle $C =$ angle C'. Now clearly $A' + B + C' =$ a straight angle. Hence, by substituting *A* for A' and *C* for C', we have $A + B + C =$ a straight angle, which is what we set out to prove.



We can't use this same proof on the sphere because it has no parallels at all. On the sphere every two straight-line paths, or great circle routes, eventually cross. In fact, on the sphere the sum of the angles of a spherical triangle is always greater than a straight angle.

On the other hand, in the plane of hyperbolic geometry there are "too many" parallels, and when the line *l* is drawn at *B* so that angle $A' =$ angle A, then angle *C* is always less than angle C'. Hence, the angle sum $A + B + C$ in triangle *ABC* is always less than the straight angle $A' + B + C'$.

Showing that an example satisfies the axioms involves interpreting the terms of the axioms as concrete entities in the particular example.

For elliptic geometry, we have as a natural model the surface of the sphere. The elliptic points of the geometry are points on the sphere and the elliptic lines are its great circles. Since the points and great circles on the sphere behave according to postulates 2–4 and E, the surface of the sphere is a model of the postulates.

To get a good understanding of hyperbolic geometry, you need to look at and ponder various models. We present several intriguing models, each with its own interpretation of "points" and "lines." First we give a simple example of a *finite* hyperbolic geometry and then consider other models.

EXAMPLE ▶ ***A Geometry of Political Alliances***

Suppose that a country has five political parties, which we will designate by *A* (Anarchists), *B* (Businessmen), *C* (Conservatives), *D* (Democrats), and *E* (Environmentalists). There are ten possible two-party political alliances: *AB*, *AC*, *AD*, *AE*, *BC*, *BD*, *BE*, *CD*, *CE*, and *DE*.

Consider the individual parties to be "points" and the two-party alliances to be "lines." We say that a "point" (party) is on a "line" (alliance) if the party is part of that alliance; for example, *A* is on *AB*. We say that two "lines" (alliances) intersect if they have a party in common; for example, *AB* and *AC* intersect in *A*. Moreover, two "lines" (alliances) are parallel if they do not have a "point" (party) in common; for example, *AB* and *CD* are parallel.

This is a small geometry indeed! It has 5 points and 10 lines. It satisfies postulate 1 of Euclid ("two points determine a line"), but not postulate 2 ("a line segment can always be extended"), and there is no mention of the circles and right angles of postulates 3 and 4. In terms of parallels, it is not a Euclidean geometry but a hyperbolic one; it satisfies postulate H, in a stronger form: If *l* is any line and *P* is any point not on the line, then there exist *exactly two* lines through *P* not meeting *l*. For example, if the line is *AB* and the point is *C*, then *CD* and *CE* are lines through *C* that are parallel to *AB*.

This geometry can be represented in terms of points and line segments, as in Figure 19.7. The parties are the vertices of a pentagon, and the alliances are line segments joining vertices. In making this interpretation, we disregard what appear as intersections of the diagonals of the pentagon; they are not points of the geometry and would not even appear if we "drew" the figure in higher-dimensional space. ◆

You can get a better understanding of hyperbolic geometry by examining models that conform to the first four postulates of Euclid and differ from ordinary Euclidean geometry only in satisfying postulate H rather than Euclid's parallel postulate (see Spotlight 19.3).

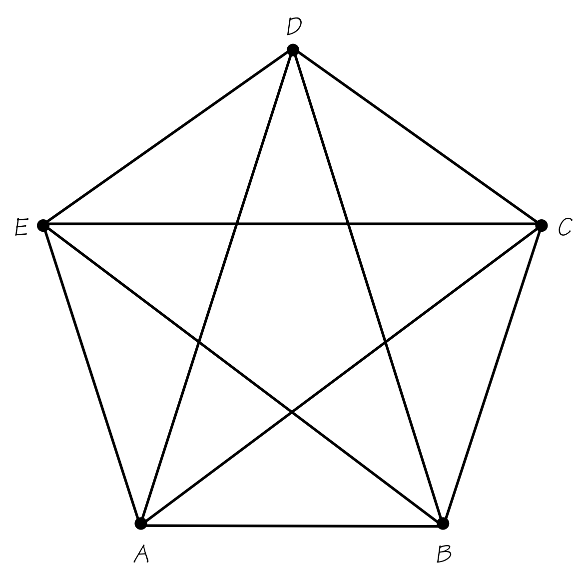
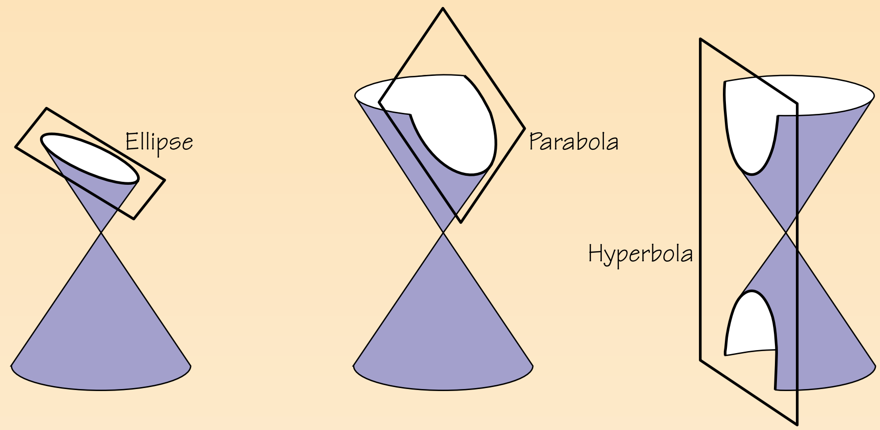


FIGURE 19.7 Representation of a geometry of political alliances.

Spotlight 19.3 What's Hyperbolic About Hyperbolic Geometry?



The word "hyperbola" comes from Greek, meaning "a throwing beyond," or excess (we get the slang word "hyper" from the same root). The word "ellipse" means "a falling short," or defect. The word "parabola" means "falling beside," or being parallel to. These terms are comparing the angle at which the conic section (hyperbola, ellipse, or parabola) cuts the cone, compared to the angle of the cone itself (see the accompanying figure).

How do these terms come to be applied to geometries? Compared to Euclidean geometry, hyperbolic geometry has an "excess" of parallels, since given a line and a point not on the line, more than one line through the point is parallel to the given line. Similarly, in an elliptic geometry, there is a "defect" of parallels, since there are no parallel lines.

In a deeper sense, the distance measure in a hyperbolic geometry has an algebraic form similar to the algebraic form of a hyperbola in analytic geometry coordinates ($y^2 - x^2 = 1$), while the distance measure in an elliptic geometry has a form similar to that of an ellipse ($x^2 + y^2 = 1$).

For the sum of the angles of a triangle, there is a reversal: hyperbolic geometry features triangles whose angle sums are "defective," while those in elliptic geometry are "excessive," compared to Euclidean geometry's 180°.

Euclidean geometry corresponds to a space with no curvature, elliptic geometry describes a space of constant positive curvature, and hyperbolic geometry describes a space of constant negative curvature.

Finally, when it comes to trigonometry, the relevant formulas are different for each geometry. The Pythagorean theorem of Euclidean geometry for right triangles does not hold in the other geometries, which have their own analogues. For spherical geometry, the formulas for trigonometry involve the familiar "circular" trigonometric functions (sine, cosine, tangent); in hyperbolic geometry, the related hyperbolic trigonometric functions apply (sinh, cosh, tanh—you may see keys for these on your calculator).

The models that we describe below have a physical interpretation that explains the fact that distances in these geometries, which we represent in the two dimensions of the Euclidean plane, are measured differently from Euclidean distances. Imagine a universe with a closed boundary that, for whatever reason, you can approach but never get to. The closer you seem to get, the harder it is to get even closer, as if there were a force pushing you away. The boundary might as well be infinitely far away. We can try to draw a picture of such a universe by drawing the boundary as a circle. To make the model work, the boundary must be infinitely far away from any point; so distances in this universe will not correspond to our ordinary notions of distance. What can the lines of this geometry—its geodesics—be? The Poincaré disk model provides an answer.

Software exercises on the models below make clear some of the amazingly different properties of hyperbolic geometry.

EXAMPLE ▶ ***The Poincaré Disk Model***

"Points" are again the interior points of a fixed circle. "Lines" are circular arcs that meet the bounding circle at right angles (the endpoints on the bounding circle are not points of the geometry). Figure 19.8 shows how postulate H is fulfilled: for the line l and the point P not on the line l, lines are shown that go through P but do not meet l.

The Poincaré disk model of hyperbolic geometry was used by the artist M. C. Escher in his fascinating "Circle Limit" prints (see Spotlight 19.4). ◆

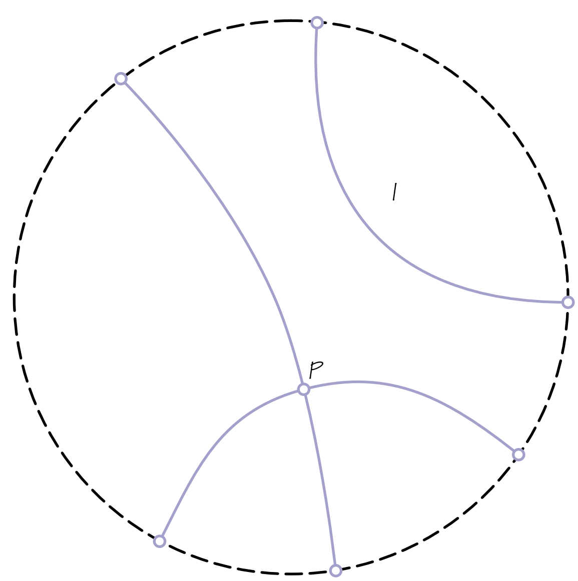


FIGURE 19.8 The Poincaré disk model of hyperbolic geometry.

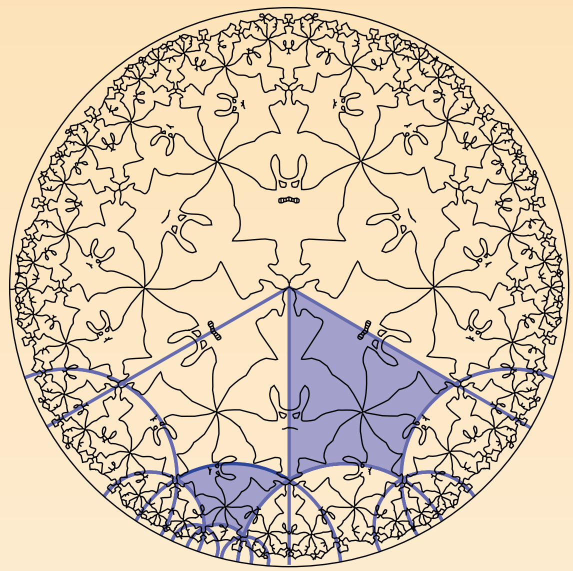
Spotlight 19.4 Angels and Devils

The Dutch artist M.C. Escher (1896–1972) was particularly interested in figures that change almost imperceptibly into other figures or into larger or smaller versions of themselves. He knew how to draw, inside a circle or square, figures that gradually get larger as they approach the outside of the enclosure. But it wasn't until he was shown a mathematician's representation of hyperbolic geometry (in which the sum of the angles of a triangle is always less than 180°) that he discovered how to make figures gradually get *smaller* toward the outside of a circle.

Douglas Dunham of the University of Minnesota, Duluth, has devised a computer program based on hyperbolic geometry that can produce an infinite variety of the type of drawings Escher so ingeniously drew. One of these, Dunham's *Circle Limit IV,* is shown here.

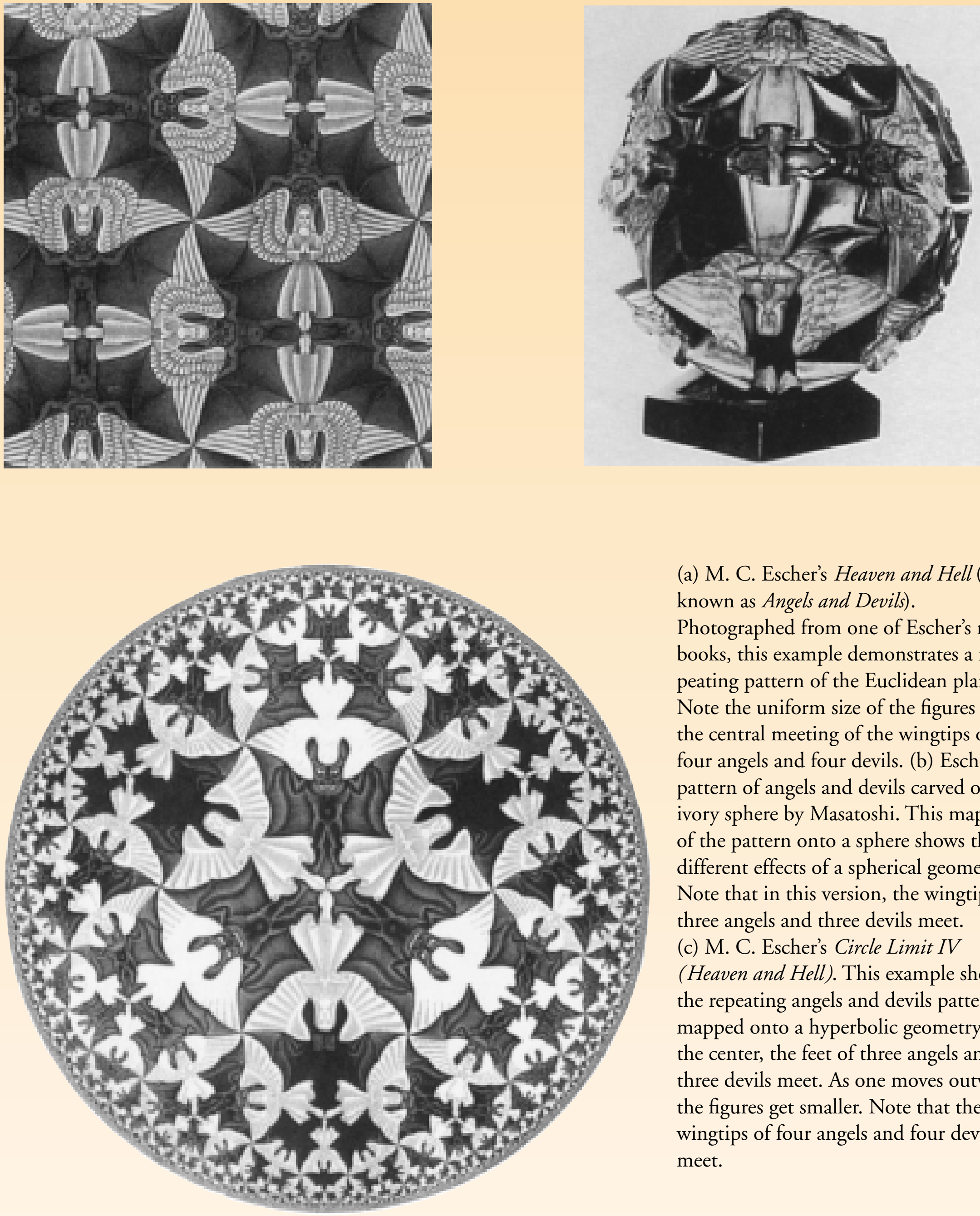
The "lines" of this geometry are arcs of circles that are perpendicular to the outside circle. (The "lines" of spherical geometry are great-circle routes on the sphere.) This particular print is based on a regular tiling of the hyperbolic plane. The tiles shown here are regular quadrilaterals—their vertices are the points where the feet of three angels meet the feet of three devils. Six of these tiles meet at each vertex. Thus, the angle of each is 60° instead of 90°, which it would be for the corresponding tiling of the Euclidean plane, where four squares meet at each vertex.

The edges of some of the tiles (of the underlying tiling) have been drawn in so that they can be seen, and two are shaded. Although the tiles appear to get smaller (in the Euclidean sense) toward the edge of the outside circle, the hyperbolic geometry uses a different distance measure in which all of the tiles, including the two that are shaded, are congruent (the same size).



Dunham's *Circle Limit IV* plot.

This computer-generated tiling in the hyperbolic plane creates an image very similar to that of M. C. Escher's *Angels and Devils.* Note how the positions of feet and heads are related by radii and arcs that define the underlying tiling pattern.



(a) M. C. Escher's *Heaven and Hell* (also known as *Angels and Devils*). Photographed from one of Escher's notebooks, this example demonstrates a repeating pattern of the Euclidean plane. Note the uniform size of the figures and the central meeting of the wingtips of four angels and four devils. (b) Escher's pattern of angels and devils carved on an ivory sphere by Masatoshi. This mapping of the pattern onto a sphere shows the different effects of a spherical geometry. Note that in this version, the wingtips of three angels and three devils meet.
(c) M. C. Escher's *Circle Limit IV (Heaven and Hell)*. This example shows the repeating angels and devils pattern mapped onto a hyperbolic geometry. At the center, the feet of three angels and three devils meet. As one moves outward, the figures get smaller. Note that the wingtips of four angels and four devils meet.

A slight variation on the Poincaré disk model produces the Klein disk model of hyperbolic geometry.

EXAMPLE ▶ ***The Klein Disk Model***

"Points" are the interior points of a fixed circle and "lines" are chords of the circle with endpoints omitted. This geometry looks more Euclidean to us who view it from outside, because the hyperbolic lines look like straight Euclidean line segments. The more unusual properties of the geometry come out when you draw figures and move them around, as the exercises ask you to do. Figure 19.9 shows an instance of postulate H being fulfilled in a particular instance, for a line l and a point P not on l. ◆

Another approach to modeling a bounded universe is to let the unapproachable closed boundary be a line that divides the plane into two halves.

EXAMPLE ▶ ***The Upper Half-Plane Model***

"Points" are the points in a half-plane excluding the bounding line. There are two kinds of "lines": open semicircles with center on the bounding line and open rays perpendicular to the bounding line. Figure 19.10 shows how postulate H is fulfilled in a particular instance. ◆

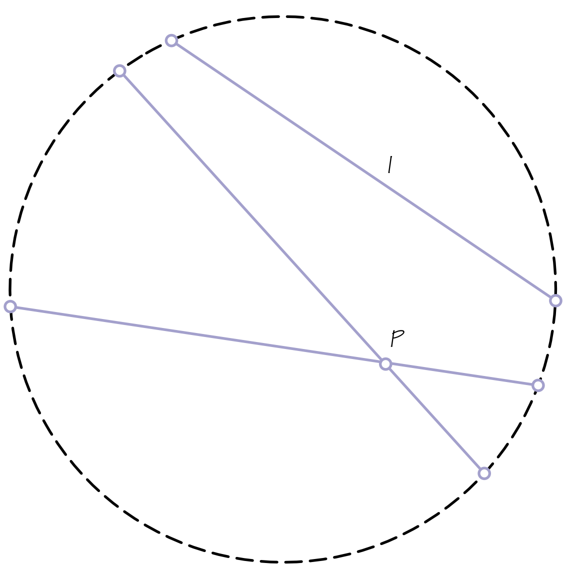


FIGURE 19.9 The Klein disk model of hyperbolic geometry.

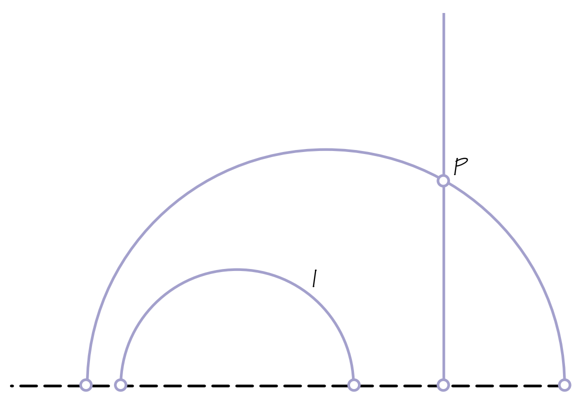


FIGURE 19.10 The upper half-plane model of hyperbolic geometry.

THE THEORY OF RELATIVITY

In 1905, Albert Einstein (Spotlight 19.5) put forth his *special theory of relativity,* a complicated theory that constituted the first step in the greatest revolution in physics since Newton's *Principia Mathematica.*

Einstein proposed a new way of thinking about events in the history of the universe. An event takes place in our three-dimensional space at a specific time in history. Thus, an event is located in *space-time* by four coordinates: three determine its position in space, and the fourth determines its position in time. Of course, these coordinates locate the event relative to a specific coordinate system. Einstein observed that the location of an event in space-time therefore depends on the position of the observer—that is, on the origin and orientation of the coordinate system being used. Different observers may obtain very different views of events, especially if one observer is traveling very fast with respect to the other.

Let's consider these ideas geometrically. The *distance* between two events, usually in relativity theory called an *interval,* is split into two parts: a *space-part* and a *time-part.* The space-part is the part of the interval that comes from the position of the events in three-dimensional space, and the time-part is the length of time that separates the events. This splitting depends on the coordinate system and its orientation, so different results may be obtained by different observers (see Figure 19.11). However, the interval, being a line segment joining the two events in four-dimensional space-time, is absolute—in the sense that it is the same for an observer at rest and for all other observers who are traveling at a constant velocity with respect to the one at rest.

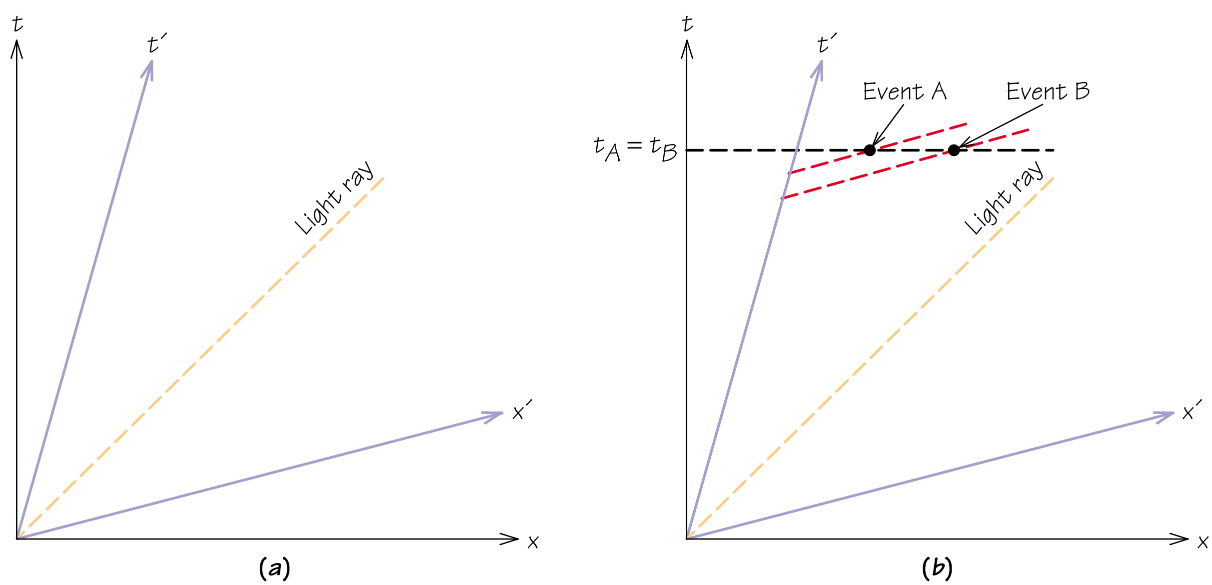
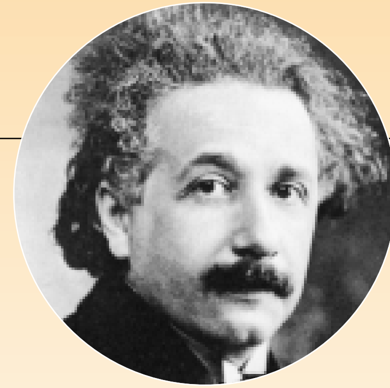


FIGURE 19.11
(a) A coordinate system representing space-time. The *t*-axes show time and the *x*-axes space. The black axes are a system at rest. Note that the blue moving system tilts toward the 45° light ray line. (b) Observers in the system at rest (black axes) will say the events *A* and *B* occur at the same time. Observers in the moving blue system will say that event *B* occurs before event *A*.

Spotlight 19.5 Albert Einstein



Albert Einstein

Albert Einstein (1879–1955) was born in Ulm to a German Jewish family with liberal ideas. Although he showed early signs of brilliance, he did not do well in school. He especially disliked German teaching methods. In the mid-1890s he went to study in Switzerland, a country much more to his liking, where he went to work as a patent clerk. Einstein burst upon the scientific scene in 1905 with his theory of special relativity. In 1916 he published his theory of general relativity. General relativity was successfully tested in 1919, and his fame grew enormously. Nazism forced Einstein to leave Europe. He settled at the Institute for Advanced Study at Princeton, where he remained until his death at age 76.

EXAMPLE ▶ ***"Simultaneous" Events***

Let's imagine that the eruption of Mount St. Helens in Washington in 1980 took place at the very same time that someone on Mount Palomar in California observed an astronomical phenomenon. Perhaps the observation was of a supernova explosion in a galaxy 100 light-years away (a **light-year** is the distance that light travels in a year). The eruption and the explosion appear to us on earth to be simultaneous.

However, knowing that the light from the supernova takes 100 years to reach earth, we realize that the eruption took place 100 years later than the supernova. For those of us on earth (and at rest relative to the earth), the interval between the two events has a space-part of 100 light-years and a time-part of 100 years.

For observers traveling at constant velocity with respect to the earth, say, at 50 light-years away from earth, the space-part and time-part of the interval would be very different. One observer might determine that the two events took place 200 years apart, while another might conclude that the two events happened simultaneously. Their splitting of the interval into space-parts and time-parts would be very different from ours. The geometry of space-time is indeed strange: in its four-dimensional space, the "distance" between two points—the interval between two events—remains invariant (in the sense we have described), but its respective parts vary. ◆

Three years after Einstein published his first paper on the subject, the mathematician Hermann Minkowski (1864–1909) gave Einstein's work a geometric interpretation that accepted Einstein's strange calculation of intervals and greatly simplified the theory. The geometry that was used, justifiably called *Minkowskian geometry,* is certainly non-Euclidean. Further, it makes use of one of Riemann's far-reaching ideas—that the nature of a mathematical space is determined by the way distance is measured; the distance formula therefore determines the nature of the geometry.

In Euclidean plane geometry, distance is determined with the help of the Pythagorean theorem $c^2 = a^2 + b^2$ about the length c of the hypotenuse of a right triangle and the lengths a and b of the legs (see Figure 19.12a), so that the distance $c = \sqrt{a^2 + b^2}$. In three dimensions, the Pythagorean theorem can be applied twice to yield the distance formula $d = \sqrt{a^2 + b^2 + c^2}$ (see Figure 19.12b); and if you imagine that in four dimensions it would be $e = \sqrt{a^2 + b^2 + c^2 + d^2}$, you would be right.

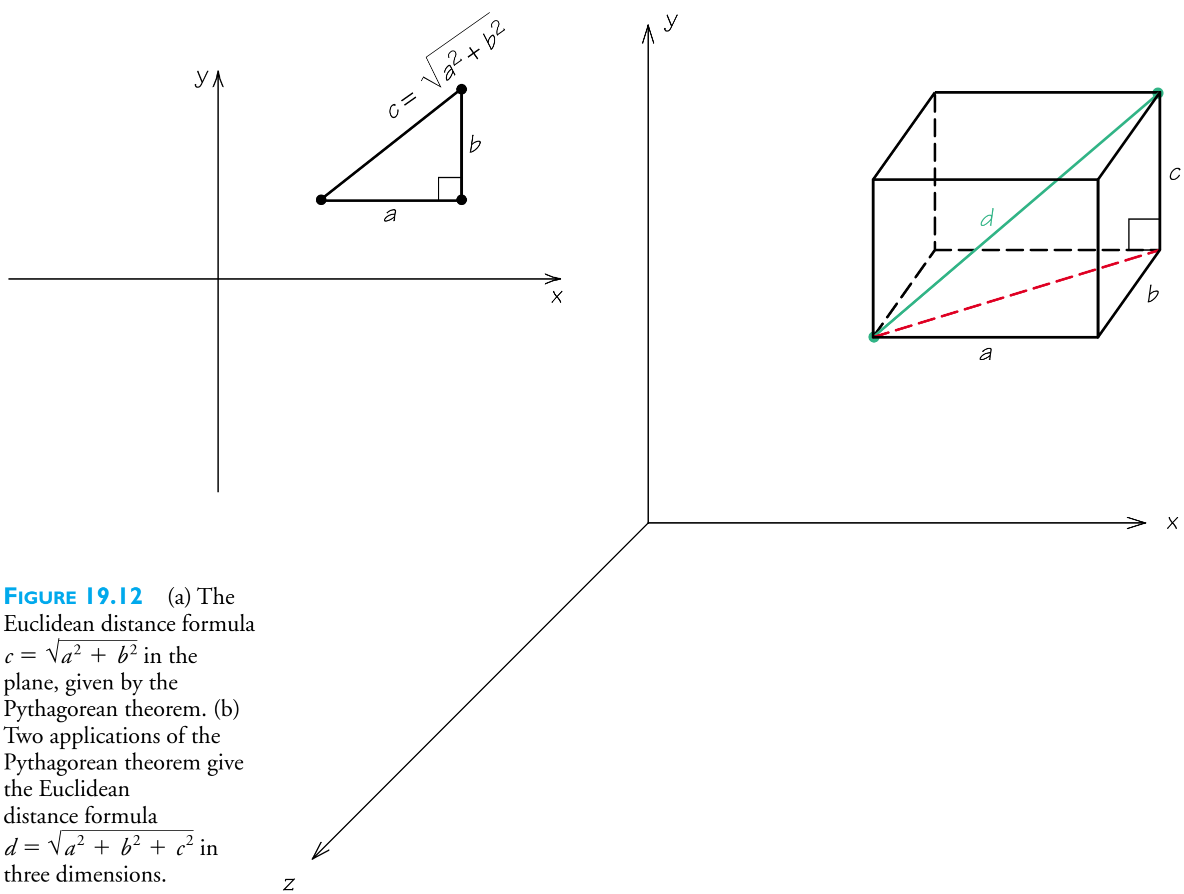


FIGURE 19.12 (a) The Euclidean distance formula $c = \sqrt{a^2 + b^2}$ in the plane, given by the Pythagorean theorem. (b) Two applications of the Pythagorean theorem give the Euclidean distance formula $d = \sqrt{a^2 + b^2 + c^2}$ in three dimensions.

If two events are A apart in the x-direction, B apart in the y-direction, and C apart in the z-direction, and T apart in time, and if the space is Euclidean four-space, then the distance d between the events would be

$$d = \sqrt{c^2T^2 + A^2 + B^2 + C^2},$$

where c is the speed of light.

In Minkowskian space, the interval I, which measures the distance between the events, is defined by

$$I = \sqrt{c^2T^2 - A^2 - B^2 - C^2}$$

This formula, with its minus signs, is very different from Euclidean distance and determines a different geometry.

General Relativity

Little more than a decade after introducing his special theory of relativity, Einstein came forth with his *general theory of relativity.* This work, too, astonished the scientific world. Among other revolutionary ideas was his contention that space was "curved." By this he meant that light rays, which are considered to travel on paths of shortest distance, don't actually follow "straight lines" but bend to follow shortest distance paths in the curved space. Light rays even bend to different degrees, depending on where in the universe they are; if they pass through a strong gravitational field, then they bend considerably.

A test of this contention was made in 1919 during a total eclipse of the sun, when the light rays from a distant star passed close to the sun and could be studied. Einstein was right; the rays did bend—and in an amount very close to his predictions. This observation showed that lines in the geometry of general relativity are not of the same character as Euclidean lines.

What sort of geometry was Einstein using? There are several answers to the question. First, the idea of "curved" space smacks of elliptic geometry, in the sense that a line through Einstein's universe comes back on itself. Second, Einstein used a variation on Minkowskian geometry in which the distance formula appropriate to the needs of physics varies from place to place in the universe, depending on the strength of the gravitational field. So, Einstein was using a form of Minkowskian geometry along with some considerably modified ideas of elliptic geometry. An appreciation of these non-Euclidean geometries very likely motivated his remark about the postulates of geometry in a famous 1921 lecture: "[They] are voluntary creations of the human mind. To this interpretation of geometry I attach great importance, for should I not have been acquainted with it, I would never have been able to develop the theory of relativity."

Relativity and Length Contraction

The road that led Einstein to relativity is marked by one crucial experiment, conducted in 1887 by A. Michelson and E. C. Morley in Cleveland at what is now Case Western Reserve University. They were trying to determine whether there was a substance, an "ether," that served as the medium for the transmission of light and electromagnetic radiation. Sound waves do not travel in a vacuum but require a medium (e.g., air or water), relative to which we can measure the speed of sound. Nineteenth-century scientists reasoned that light waves, which do pass through the vacuum of space, must be carried through it in some medium that had not yet been detected.

The essence of their experiment was to send a lightbeam out to a mirror and back and measure the time that elapsed. They did this for two situations. In one, the lightbeam was aligned in the direction of the earth's rotation on its axis, so that on the way out the speed of the earth would add to the speed of light through the ether, and on the way back, it would subtract from the speed of the light. In the other situation, they aligned the lightbeam perpendicular to the earth's rotation on its axis, so that the earth's spinning would not affect the *speed* of the lightbeam in the ether. In this situation, however, the earth's rotation does affect *how far* the lightbeam has to travel.

What Michelson and Morley expected to observe was that the speed of light would be different for the two situations.

EXAMPLE ▶ ***Swimming in a Current***

We imagine an analogous situation. You are in a river that is a distance d wide and you decide to conduct a swimming experiment. In place of bouncing lightbeams, you will swim back and forth. Your speed of swimming c corresponds to the speed of light, and the speed v of flow of the river corresponds to the speed of the rotation of the earth.

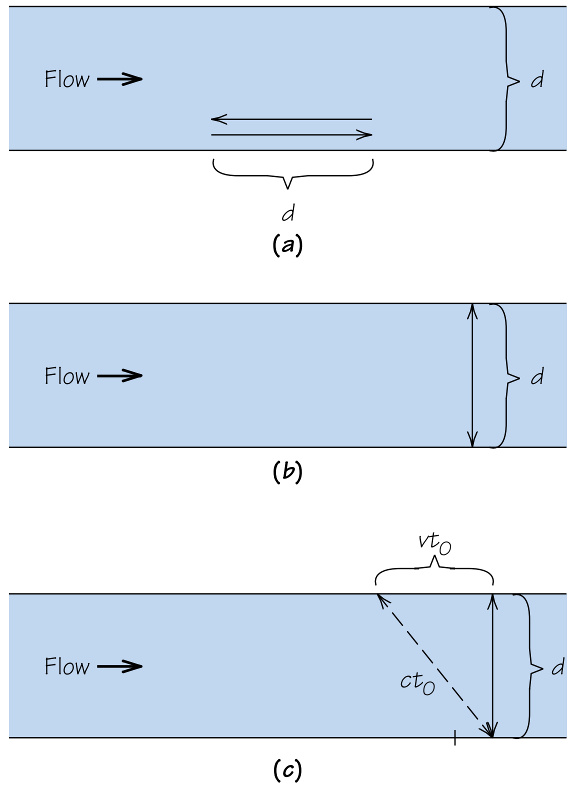
Corresponding to the first situation, you swim a fixed distance d downstream, then swim back again (see Figure 19.13a); let's say that it takes you time t_1 downstream and time t_2 upstream. On the trip downstream, you are swimming through the water at speed c but you are moving at a speed of $(c + v)$ relative to the bank. Similarly, on the trip upstream, you are really moving only at speed $(c - v)$.

Using the basic formula distance = rate × time, we have $d = (c + v)t_1$ and $d = (c - v)t_2$. Your total time for the round-trip is

$$t_{\text{up}+\text{down}} = t_1 + t_2 = \frac{d}{c+v} + \frac{d}{c-v}$$

$$= \frac{2dc}{(c+v)(c-v)} = \frac{2d/c}{1 - v^2/c^2}$$

Figure 19.13
(a) Swimming with and against the current of a river. (b) Swimming across a river and back to the starting point. (c) To stay even with the starting point, the swimmer needs to head upstream to some degree.



You should check these calculations for a numerical example; for example, the distance d is 1/2 mile, you can swim at a speed c of 2 miles an hour, and the river flows at a speed v of 1 mile per hour. You should find that it takes you one-sixth of an hour downstream and one-half of an hour upstream, for a total time for the round-trip of two-thirds of an hour, or about 0.67 hr.

Corresponding to the second situation, you swim distance d across the river and back to your starting point (see Figure 19.13b). As you swim, however, you have to fight the downstream current, so that you always stay even with your starting point. You not only have to swim the distance across the river, you also need to swim upstream the equivalent distance that the river is carrying you downstream; you need to always head part way upstream, so only part of your swimming effort is directed toward carrying you directly across the river. In effect, you are swimming the length of the hypotenuse of a right triangle, as shown in Figure 19.13c, and being carried downstream by the length of one leg.

Denote the time that it takes you to cross the river by t_0. The relationship between how far you travel and how long it takes is given by the Pythagorean theorem, as mentioned earlier:

$$(ct_0)^2 = vt_0^2 + d^2$$

We solve the equation for the time:

$$c^2t_0^2 - v^2t_0^2 = d^2,$$

$$t_0^2 = \frac{d^2}{c^2 - v^2},$$

$$t_0 = \frac{d}{c\sqrt{1 - v^2/c^2}}$$

The time for the return trip is the same, so the total time for the round-trip is

$$t_{\text{back+forth}} = \frac{2d/c}{\sqrt{1 - v^2/c^2}}$$

Testing this algebra with the same numerical values as before, you should find that the round-trip across the river takes you $1/\sqrt{3} \approx 0.577$ hr.

You observe that the times for the two round-trips are different. The time for the round trip in the direction of motion of the current and back is longer, by a factor of $1/\sqrt{1 - v^2/c^2}$, than the trip across the current and back. ◆

Michelson and Morley had a different experience: the times that they observed in their analogous experiment were *the same.* Because of their experience and reputation as experimenters, they and their colleagues were sure that this astonishing result was not the consequence of measurement or experimental error.

So what was wrong with the theory? One ingenious explanation is that the speed of the rotation of the earth somehow *contracts lengths along the direction of motion,* in exactly the right proportion to cancel the difference in times. In our swimming example, it would be as if distance up and down stream were made shorter by the current. Similarly, a planet moving at nearly the speed of light would flatten in the direction of motion, so as to turn into a pancake. For this explanation to work, the length l' along the direction of motion would have to be related to the length l in the direction perpendicular to the motion by

$$l' = l\sqrt{1 - v^2/c^2} = l/\gamma$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

is known as the **Lorentz-Fitzgerald factor.**

Because the speed of light is so great—186,000 miles per second, or 3.0×10^{10} cm/sec—the value of the Lorentz-Fitzgerald factor is very nearly 1 until v reaches about 10% of the speed of light.

Why couldn't Michelson and Morley detect this shrinkage in the direction of motion? Because their measuring rods, when aligned along the direction of motion, shrank too. So the contraction theory could never be verified by direct measurements.

You would think that photography would help—for example, that you could see a ball moving at nearly the speed of light as the pancake shape that it must be. However, most surprisingly, even a camera can't see the contraction! An optical distortion compensates for the shrinkage, as we now explain.

You and the camera see by means of "particles" (photons) of light reflected from an object. Light from faraway objects can take a long time to reach us: light from the sun takes about eight minutes, and the light that reaches us now from distant stars was emitted billions of years ago. Similarly, in the case of a moving object, we see at the same time images of close-up parts and (because of time delay from the fixed speed of light) earlier images of faraway parts of an object. Hence, the object appears (to us or on the film) stretched in the direction of motion. This stretching compensates for the contraction. Figure 19.14 shows a computer reconstruction of how an object moving at close to the speed of light would appear. Notice that it does not appear to shrink in the direction of motion (even though it actually does).

The Lorentz-Fitzgerald contraction theory was based on a complicated theory of matter interacting with the ether. Scientists eventually were forced to conclude that there is no medium in which light waves move, no "ether" relative to which we can measure the speed of light. The appealing analogy between a light wave and a swimmer swimming through water is misleading.

Twenty-four years after the Michelson-Morley experiment, Einstein hypothesized that the speed of light is not affected by the motion of the source nor of the observer. The rotation of the earth cannot add to or subtract from the speed of light in the Michelson-Morley experiment. Einstein's theory predicts the same time, $2d/c$, for the round-trip in either alignment of their equipment.

Einstein's theory of relativity also predicts a contraction of length in the direction of relative motion, by exactly the Lorentz-Fitzgerald factor. The reason for this relativistic contraction, however, has nothing to do with ether or with Lorentz's theory to explain the Michelson-Morley results on the basis of it.

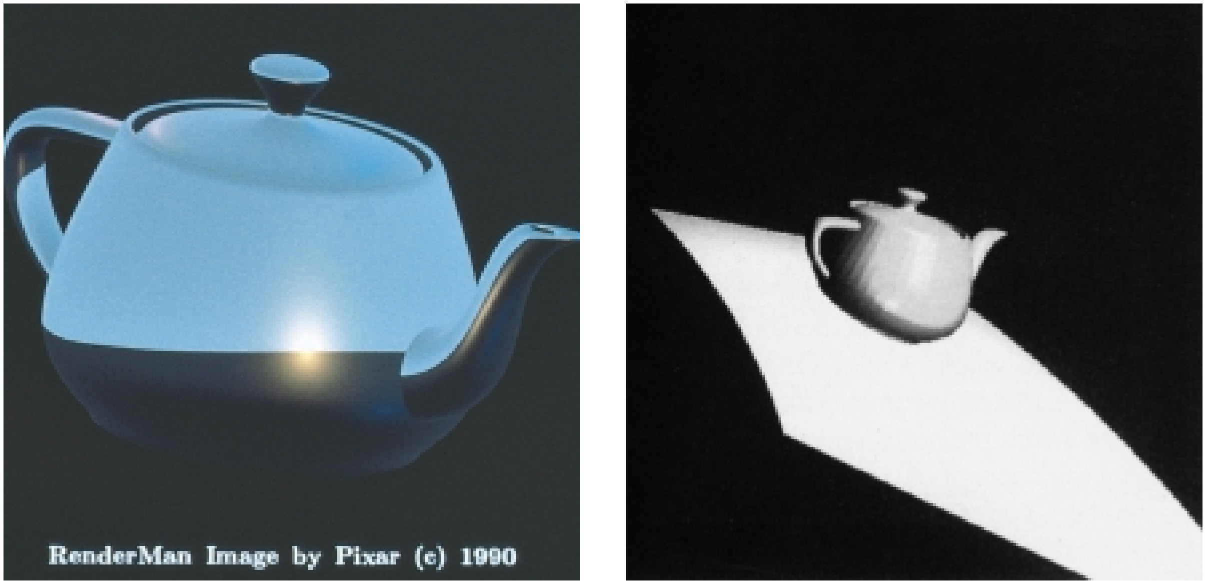


FIGURE 19.14 (a) A teapot at rest. (b) A view from the same angle of the teapot passing by at 99% of the speed of light.

Einstein's theory eliminates the need to suppose that there is an ether. The underlying reason for the relativistic length contraction is relativity itself: the relative motion of object and observer. As viewed by you, the teapot is moving at close to the speed of light, and its length must contract in the direction that it appears to you to be traveling (even though you can't see the contraction, as we've explained). As viewed by an ant on the teapot, *you* appear to be moving at close to the speed of light and are thin as a pancake in the direction that you appear to be moving (even though the ant can't see you that way).

Another consequence of relativity is that *time, too, contracts with motion.* Consider two observers moving at a constant velocity relative to each other. Each will observe that the other's clock runs slow compared to their own, by a factor of γ. This strange result is known as the *clock paradox.*

WHICH GEOMETRY IS TRUE?

As far as measurement and travel on the surface of the earth go, we know that we live on a world with an elliptic geometry. When it comes to travel at velocities near the speed of light, the geometry that applies to space-time is Minkowskian geometry, a kind of non-Euclidean geometry. But what about the universe of space beyond the earth's surface, without regard to time? Do we really live in a spatial universe that is Euclidean?

Spotlight 19.6 The Implications of Non-Euclidean Geometry and Relativity

The creation of non-Euclidean geometry affected scientific thought in two ways. First of all, the major facts of mathematics, that is, the axioms and theorems about triangles, squares, circles, and other common figures are used repeatedly in scientific work. Since these facts could no longer be regarded as truths, all conclusions of science that depended upon strictly mathematical theorems also ceased to be truths.

Second, the debacle in mathematics led scientists to question whether they could ever hope to find a true scientific theory.

Even on the level of engineering, a serious question emerged. Since bridges, buildings, dams, and other works were based on Euclidean geometry, was there not some danger that these structures would collapse? But this thought did not alarm the scientists and engineers of the nineteenth century, who did not believe that the geometry of physical space could be other than Euclidean. However, the advent of the theory of relativity drove home the point that Euclidean geometry is not necessarily the best geometry for applications. For engineering involving motion with high velocities, such as modern accelerators of electrons or neutrons, the theory of relativity is used.

Past ages have sought absolute standards in law, ethics, government, economics, and other fields. They believed that by reasoning one could determine the perfect state, the perfect economic system, the ideals of human behavior, and the like. This belief in absolutes was based on the conviction that there were truths in the respective spheres. But in depriving mathematics of its claim to truth, the non-Euclidean geometries shattered the hope of ever attaining any truths.

The view that mathematics is a body of truths was accepted at face value by every thinking being for 2000 years. This view, of course, proved to be wrong. We see, therefore, on the one hand, how powerless the mind is to recognize the assumptions it makes. Apparently we should constantly reexamine our firmest convictions, for these are most likely to be suspect. They mark our limitations rather than our positive accomplishments. On the other hand, non-Euclidean geometry also shows the heights to which the human mind can rise. In pursuing the concept of a new geometry, it defied intuition, common sense, experience, and the most firmly entrenched philosophical doctrines just to see what reasoning would produce.

Source: Adapted from Morris Kline, *Mathematics for the Nonmathematician,* Dover, New York, 1985, pp. 474–476.

One way to tell would be to measure the angles in a large triangle, to see how their sum compares with 180°. Because all measurements have some imprecision, we cannot prove that the angles of a measured triangle add up to exactly 180°. Sufficiently precise measurements of very large triangles, however, could conceivably prove that space is not Euclidean.

Gauss was employed for a time by the government of Hanover in a geodetic survey, in the course of which he measured the angles in a triangle formed by three mountain peaks roughly 50 miles apart. The deviation from

180° was less than the error estimate for the measurement, so the sum could be equal to 180°, or greater, or less—the sum was consistent with all three hypotheses. In fact, if there is any difference from 180° for this triangle of mountain peaks, it was far too small for Gauss to detect—as he probably realized—and far too small even for us to detect today.

Lobachevsky considered even larger triangles and looked into the parallax of stars (the apparent relative motion, as the earth orbits the sun, of nearer stars compared to more distant ones). But neither he nor others since have found a triangle whose angle sum is definitely different from 180°, despite the fact that in hyperbolic geometry, the larger the area of the triangle, the larger the defect must be.

If space does have a hyperbolic geometry, then there is a lower bound for the parallax of stars (though that would not mean that there is a limit to how far away stars can be). Although there is certainly a smallest observed parallax among the thousand or so stars whose parallax we know, there may be stars yet unmeasured whose parallax is even smaller.

Space could have an elliptic geometry, with triangles having angular excesses rather than defects. The universe could then be the three-dimensional analogue of the two-dimensional surface of a sphere. Just as the surface of a sphere has a finite area, the universe could have a finite volume, despite having no boundaries—just as the surface of a sphere has no boundaries. Such a space would have positive curvature.

Regardless of the true situation for actual three-dimensional space, we appear to perceive space visually as hyperbolic. Common visual illusions, classic experiments in perception, and the empirical truth of *Brentano's hypothesis* (that humans tend to overestimate small angles and underestimate large ones) all lead to the conclusion that "perceived space" is hyperbolic.

The question of which geometry is true would not have occurred to anyone before the nineteenth century. The discovery of non-Euclidean geometry and the theory of relativity have had profound intellectual implications in all fields (see Spotlight 19.6).

REVIEW VOCABULARY

Elliptic geometry A geometry in which there are no parallel lines.

Euclidean geometry The "ordinary" system of geometry based on the five postulates Euclid used, including Euclid's parallel postulate.

Euclid's parallel postulate If l is any line and P any point not on l, then there exists exactly one line through P that does not meet l—in other words, given a line and a point not on the line, there is exactly one other line that passes through the point and is also parallel to the given line.

Geodesic A curve (possibly straight) that gives the shortest path between two points on a surface.
Great circle The set of points that is the intersection of a sphere and a plane containing its center.
Hyperbolic geometry A geometry in which for any given line and a point not on the line, more than one line both passes through the point and is parallel to the given line.
Light-year The distance that light travels in a year.
Logically equivalent Two statements are logically equivalent if each can be deduced from the other.
Logically independent A statement is logically independent of a collection of other statements if neither it nor its negation can be deduced from them.
Lorentz-Fitzgerald factor The factor by which length and time are contracted by motion.
Model An example that satisfies a collection of axioms is a model of the axioms.
Non-Euclidean geometry Any collection of postulates, theorems, and corollaries for geometry (concerning points, lines, circles, and angles) that differs from the collection formulated by Euclid.
Parallel lines Two lines are parallel if they do not meet.
Parallel postulate A basic assumption of geometry that states whether through a point P, not on a given line l, there exists none, one, or more than one line parallel to given line l.
Postulate E Every two lines intersect.
Postulate H If l is any line and P is any point not on the line, then there exists more than one line through P not meeting l.
Spherical geometry The geometry of a sphere or the earth's surface.

SUGGESTED READINGS

ABBOTT, EDWIN A. *Flatland: A Romance of Many Dimensions,* Princeton University Press, Princeton, N.J., 1989.

BOLOTOVSKY, B. M. What's that you see? On the perceived shape of rapidly moving objects, *Quantum,* 3 (4) (March/April 1993): 5–8.

BOLTYANSKY, VLADIMIR. Turning the incredible into the obvious: How many geometries do you know? *Quantum,* 3 (1) (September/October 1992): 19–23.

BURGER, DIONYS. *Sphereland: A Fantasy About Curved Spaces and an Expanding Universe,* Crowell, New York, 1965. Reprinted with a new introduction, Harper & Row, New York, 1983.

CARVER, MAXWELL. Brain bogglers: Chicken a la king, *Discover* (March 1988): 96, 92.

CROWE, DONALD W. Some exotic geometries, in Anatole Beck, Michael Bleicher, and Donald W. Crowe, eds., *Excursions into Mathematics,* Worth, New York, 1969, pp. 211–314. A comparison of spherical, Euclidean, and hyperbolic geometries, with sections on finite geometries.

DAVIS, A. S. The relevance of mathematics for the remnant, University of Oklahoma, Department of Mathematics Preprints No. 48, 1968.

DUBROVSKY, VLADIMIR. Inversion: A most useful kind of transformation, *Quantum,* 3 (1) (September/October 1992): 40–46; and (November/December 1992): 80–81.

GINDIKIN, SIMON. The wonderland of Poincaria, *Quantum,* 3 (2) (November/December 1992): 21–26, 58–59.

GREENBERG, M. J. *Euclidean and Non-Euclidean Geometries,* Freeman, New York, 1980.

HSIUNG, PING-KANG, ROBERT H. THIBADEAU, AND ROBERT H. P. DUNN. Ray-tracing relativity, *Pixel,* 1 (1) (January/February 1990): 10–18.

JACOBS, HAROLD R. *Geometry,* 2nd ed., Freeman, New York, 1987. Chapters 6 and 16 treat parallel lines and non-Euclidean geometries.

MAURER, STEPHEN B. The king chicken theorems, *Mathematics Magazine,* 53 (1980): 67–80.

PARKER, GEORGE D. *Poincaré: An Excursion into Hyperbolic Geometry,* IBM-PC program for doing hyperbolic geometry in the Poincaré upper-half-plane model. Available from the author at 1702 West Taylor, Carbondale, IL 62901.

PETERSON, IVARS. Space-time odyssey: Visualizing the effects of traveling near the speed of light, *Science News,* 137 (April 14, 1990), 232–233, 237.

PETIT, JEAN-PIERRE. *Everything Is Relative: The Adventures of Archibald Higgins,* William Kaufmann, Inc. Relativity as told in a cartoon tale, "for adults of any age."

SMART, JAMES R. *Modern Geometries,* 4th ed., Brooks/Cole, Pacific Grove, Calif., 1994.

SVED, MARTA. *Journey into Geometries,* Mathematical Association of America, Washington, D.C., 1991. Introduces hyperbolic geometry in a delightfully informal style.

TRUDEAU, RICHARD J. *The Non-Euclidean Revolution,* Birkhäuser, Boston, 1987.

WEEKS, JEFFREY R. *The Shape of Space: How to Visualize Surfaces and Three-Dimensional Manifolds,* Dekker, New York, 1985. Contains chapters on the hyperbolic plane (including how to make hyperbolic paper), spherical geometry, hyperbolic space, and three-dimensional spherical geometry.

ZAGE, WAYNE M. The geometry of binocular visual space. *Mathematics Magazine,* 53 (1980): 289–294.

EXERCISES

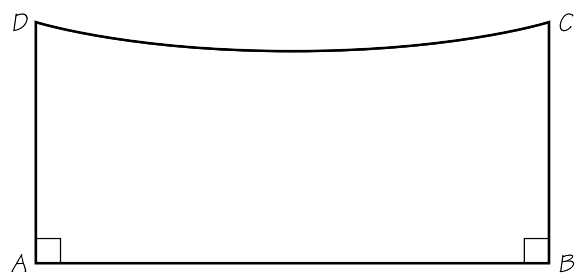
▲ *Optional.* ■ *Advanced.* ◆ *Discussion.*

Euclidean Geometry

1. In triangle ABC in the Euclidean plane, the measure of $\angle A$ is twice the measure of $\angle B$, and the measure of $\angle C$ is three times the measure of $\angle B$. Determine all three measures.

Hyperbolic Geometry

For Exercises 2 and 3, refer to the following: You are working in hyperbolic geometry, and the quadrilateral $ABCD$ in the following figure is drawn to suggest the situation of a quadrilateral in hyperbolic geometry with right angles at A and B and acute angles at C and D.



2. Prove that either the sum of the angles of triangle ABD is less than 180°, or the sum of the angles of triangle BCD is less than 180°, or both. (*Hint:* What can you say about the sum of the angles of the quadrilateral?) (In fact, both triangles have angle sum less than 180°, but that is harder to prove.)

3. Prove that if $AD = BC$, then angles C and D are of equal measure. (*Hint:* Draw the diagonals. Then use the Euclidean theorems on the congruence of triangles, which do not depend on Euclid's parallel postulate.)

Elliptic Geometry

4. By producing a specific example, show that there is a triangle in elliptic geometry in which all three angles are right angles, so that the sum of the angles of the triangle is 270°. (*Hint:* Spherical geometry is an elliptic geometry.)

Models for Hyperbolic Geometry

Exercises 5–10 require a Macintosh computer and the freeware program SnapPeaFunAndGames.sea.hqx

- from the World Wide Web source http://archives.math.utk.edu/software/mac/geometry/.directory.html, or
- by anonymous ftp transfer from softlib.rice.edu in the directory pub/NonEuclid, or
- by using a Gopher client to archives.math.utk.edu, proceeding to Software (Packages, Abstracts and Reviews), to Macintosh Software arranged by subjects, to Geometry.

From any of these sources, you must use BinHex 4.0, StuffIt Expander, or similar helper program to convert the downloaded file into application programs. The program that you are interested in is hyperbolic MacDraw. After launching the program, you will be presented with a window entitled "Poincaré Disk Model" containing a circle. Within this circle you can use the mouse to draw lines of hyperbolic geometry, including lines that extend to the boundary of the circle. Other tools in the Tool menu allow you to draw closed figures, such as triangles, quadrilaterals (four-sided polygons), and regular polygons (all sides the same length) with any number of sides. Changing tools to the hand tool, you can move figures around inside hyperbolic space. Moreover, the Model menu gives you the option to also view your figures in the Klein disk model and in the upper half-plane model.

5. When you first launch the program, you are in the Poincaré disk model and the tool for drawing a line is active. Start with the pen cursor on the boundary of the circle, hold the mouse button down, and move the cursor to elsewhere on the boundary. Notice that wherever you end the line, both ends of the line make a 90° angle with the boundary.

6. The program does not provide for erasing figures. Under File select New to get a new hyperbolic plane. Start with the pen cursor inside the circle, hold the mouse button down, and move to another point inside the circle, passing through the center of the circle. You have drawn a line segment, which should look fairly straight to you. Change to the hand tool and move the segment around. Notice that as you move the segment it appears to curve. Also, as you move the segment closer to the boundary, it appears to get shorter. That is your perspective from viewing the hyperbolic plane from the outside; from inside, the segment moves around, but always lies along some line of the geometry (all lines make 90° angles with the boundary) and remains the same length.

7. Again get a new hyperbolic plane. Use the tool in the lower left-hand corner of the Tool menu to draw a triangle, which the program will automatically shade. Use the hand tool to move the triangle around. As you move it, its

sides appear to change shape and it appears to change area. (Careful! If you move it too close to the boundary, you can lose it!) Again, that is a perspective from outside the geometry; from inside, the sides always lie along lines and the area remains the same.

8. Get a new hyperbolic plane and select the polygon tool on the right-hand side of the Tool menu. Click in the exact center of the circle. The program will prompt you for the number of sides for the polygon (enter 8) and for the angle between two adjacent sides (enter 4 to get $2\pi/4$; 2π radians corresponds to 360°, so the sides of this polygon will be at 90° to each other). What happens as you move this octagon around?

9. As in part (d), but let the polygon have four equal sides. Try various values for the divisor for 2π. Which values give you something that when positioned in the middle of the hyperbolic plane looks most like a square? What are the angles between its adjacent sides? What is the smallest divisor for 2π that will produce a hyperbolic polygon with four equal sides, and what is the angle between adjacent sides?

10. Take each of the hyperbolic planes in which you have drawn figures in Exercises 5–9 and convert first to the Klein disk model and then to the upper half-plane model. What can the sides of a hyperbolic polygon with four equal sides look like in the latter model?

Exercises 11–16 require a Macintosh computer. The Macintosh program NonEuclid is available without fee for educational use from the sources below. (Similar and additional capabilities are provided by the MSDOS program Poincaré for doing hyperbolic geometry in the upper half-plane model instead of the Poincaré disk model. That program is available for a fee from the author, George D. Parker, 1702 West Taylor, Carbondale, IL 62901.) NonEuclid can be obtained

- from either of the World Wide Web sources http://archives.math.utk.edu/software/mac/geometry/.directory.html, or http://riceinfo.rice.edu:80/projects/NonEuclid/NonEuclid.html, or
- by anonymous ftp transfer from softlib.rice.edu in the directory pub/NonEuclid, or
- by using a Gopher client to archives.math.utk.edu, proceeding to Software (Packages, Abstracts and Reviews), to Macintosh Software arranged by subjects, to Geometry.

From any of these sources, you must use BinHex 4.0, StuffIt Expander, or similar helper program to convert the downloaded file into the application program and associated files. Unlike hyperbolic MacDraw, this program does not allow you to move figures once you have constructed them; but it does do measurements for you of lengths, angles, and areas of figures.

11. Under Help, read Introduction and My First Triangle. Follow the instructions of My First Triangle to draw a triangle, measure its angles, and find its angle sum.

12. Most of the remaining options under Help ask you to use the program to perform constructions to test whether theorems of Euclidean geometry are true or not in hyperbolic geometry. Read the text under the corresponding title and do constructions to try to answer the questions there about whether theorems of Euclidean geometry carry over. Begin with What To Do, which tells how to go about your experiments, and then investigate

(a) parallel lines
(b) circles
(c) angles

13. Repeat Exercise 12, but for

(a) parallelograms
(b) rhombuses

14. Repeat Exercise 12, but for

(a) rectangles
(b) squares

15. Repeat Exercise 12, but for

(a) triangles
(b) isosceles triangles
(c) equilateral triangles
(d) right triangles

16. Use Open under File to open and explore files in the folder NonEuclid Examples, which discuss area, equilateral triangles, radii of circles, tilings by triangles, and whether a coordinate system is possible in hyperbolic geometry. The tiling of Web of Congruence is the pattern for M. C. Escher's print *Circle Limit IV (Heaven and Hell)* in Spotlight 19.4 (page 772).

For Exercises 17–24, refer to the following: A collection of trees is arranged in rows so that the following axioms are satisfied.

A There is at least one tree.
B Each row contains exactly two trees.
C Each tree belongs to at least one row.
D Any two trees have exactly one row in common.
E For any row, there is exactly one other row with no trees in common with the first row.

17. Show the following:

    (a) There is at least one row.
    (b) There are at least two rows.
    (c) There are at least four trees.

18. Show the following:

    (a) Every tree belongs to at least two rows.
    (b) There are at least six rows.

19. Show the following:

    (a) There are exactly four trees.
    (b) There are exactly six rows.
    (c) Each tree belongs to exactly three rows.

20. In this problem, you create models of the axiom system by interpreting the terms used in the axioms:

    (a) Interpreting "tree" as a point and "row" as a line, draw a model of the axiom system. Note that the lines that you draw in the plane have many other planar points that belong to them but that we don't consider, as they are not part of this particular axiom system. (This gives a *four-point geometry.*)
    (b) Interpreting "tree" as a line and "row" as a point, draw a model of the axiom system. (This gives a *six-point geometry.*)

21. In this problem, you create further models of the axiom system:

    (a) Interpreting "tree" as student and "row" as committee, construct a model of the axiom system. Name or label the students and list all of the members of each committee.
    (b) Interpreting "tree" as committee and "row" as student, construct a model of the axiom system. Name or label the students and list all of the members of each committee.

22. Interpreting "tree" as a line and "row" as a point:

    (a) How many triangles are there? (A triangle consists of three points and three lines joining them in pairs.)
    (b) Are there two lines that are parallel? (Two lines are parallel if they do not have a point in common.)
    (c) Does Euclid's parallel postulate hold?

23. Interpreting "tree" as a point and "row" as a line:

    (a) How many triangles are there? (A triangle consists of three points and three lines joining them in pairs.)

(b) Are there two lines that are parallel? (Two lines are parallel if they do not have a point in common.)
(c) Does Euclid's parallel postulate hold?

24. [Adapted from Crowe (1969), p. 289.] Suppose that "point" means the location of a store where groceries are sold — say, Albany, Birmingham, Chicago, or Denver. Suppose that "line" means a particular line of groceries — say, apples, bananas, cheese, doughnuts, eels, or figs. Finally, suppose that "a tree belongs to a row" or "a row contains a tree" means that a particular line of goods is sold at a particular location. Give an explicit distribution of the six commodities in the four cities so that the axioms apply.

For Exercises 25–28, refer to the following: Suppose the following set of axioms concerns two classes of objects K and L, whose nature is left undetermined:

A Any two members of K are contained in exactly one member of L.
B No member of K is contained in more than two members of L.
C There is no member of L that contains all of the members of K.
D Any two members of L contain exactly one member of K in common.
E No member of L contains more than two members of K.

We are interested in possible models of these axioms. Check the axioms to verify that if neither K nor L has any members at all, the axioms still hold. We say that they are vacuously true for this "empty" model.

25. For each of the following situations, is there a model? If so, verify that the axioms all hold; if not, show what axioms would have to be violated.

(a) K has no members and L has at least one.
(b) K and L each have one member.

26. Is there a model in which:

(a) K has one member and L has two members?
(b) K has one member and L has three or more members?

27. Add the additional axiom

F K has at least two members.

(a) Show that K has at least three members.
(b) Show that K cannot have more than three members, so it has exactly three members. (This gives a *three-point geometry*.)
(c) How many members can L have?
(d) Does this axiom system satisfy Euclid's parallel postulate?

28. As in Exercise 27,

    (a) Using points and lines, give a model of the set of axioms.
    (b) If axiom C is now omitted, are other models possible?

For Exercises 29–32, refer to the following: Consider the following system of axioms concerning two classes of objects M and N.

A N has at least one member.
B Every member of N contains exactly three members of M.
C Any two members of M are contained in exactly one member of N.
D There is no member of N that contains all of the members of M.
E Any two members of N have at least one member of M in common.

29. We explore how many members M and N may have.

    (a) Show that each two members of N have exactly one member in common.
    (b) Show that each of M and N has at least seven members. To help your intuition, you may want to try to construct a model of the axioms out of points and lines. Remember, though, that you must reason from the axioms, not from any picture that you draw.

30. We continue with determining the number of members of M and N.

    (a) Suppose that M has an eighth member. Show that this supposition leads to a contradiction.
    (b) Show that M and N must have exactly seven members each.

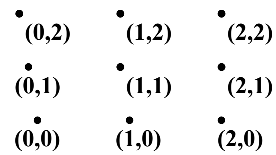
31. What are some models for this set of axioms?

    (a) Interpret the members of M as students and the members of N as committees. Name or label the students and list all of the members of each committee.
    (b) Interpret the members of M as points and the members of N as lines, and draw an appropriate model. (This is *Fano's seven-point geometry.*)

32. For the model in terms of points and lines in Exercise 30(b):

    (a) Are there two lines that are parallel? (Two lines are parallel if they do not have a point in common.)
    (b) Does Euclid's parallel postulate hold?

For Exercises 33–37, refer to the following: Consider the set {0, 1, 2} under a "clock" arithmetic, with 3 taking the role of the 12 in the usual clock arithmetic, so that $1 + 2 = 0$ and $2 + 2 = 1$. We can even introduce multiplication, with the usual results except that $2 \times 2 = 1$ (we would expect 4, which converts to 1 in our clock arithmetic). Using the set {0, 1, 2} as the possible x- and y-coordinates, we can form the nine points in the figure below. These will be the points of a "miniature" analytic geometry. The lines will consist of points that satisfy linear equations, that is, equations of the form $ax + by = c$, with a, b, and c from the set {0, 1, 2}.



33. List all of the points on the line

    (a) $x = 1$
    (b) $y = 2$
    (c) $x + y = 1$
    (d) $x + 2y = 1$

34. Find the intersection of the lines $x + y = 1$ and $2x + y = 2$

    (a) by using algebra. (*Hint:* Subtract one equation from the other.)
    (b) by listing all of the points on each line and comparing the lists.

35. We explore how many lines and points are in this geometry:

    (a) How many different lines are there? (*Hint:* A line is either a vertical line, with equation $x = c$, or else it can be written in the form $y = mx + b$, with each of c, m, and b being either 0, 1, or 2.)
    (b) How many points lie on each line?
    (c) How many lines pass through each point?

36. This miniature analytic geometry is known as the *affine two-dimensional geometry over 3 elements,* an example of an *affine plane.* It satisfies the following three axioms:

    A For any pair of distinct points, there is exactly one line containing both of them.

    E Given a line and a point not on the line, there is exactly one line through the point that does not contain any points of the given line. (This is the parallel postulate, so this is a Euclidean geometry.)

B There are at least four points, no three of which lie on the same straight line.

(a) Prove that axiom A is satisfied. (*Hint:* Use the two-point formula for the equation of a straight line,

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}$$

to exhibit one such line. Suppose that there are two distinct lines through the same two points. Take one of the points and count all of the points on the lines that pass through that point. Show that a contradiction results.)

(b) Prove that axiom E is satisfied. (*Hint:* Join the given point to every point on the given line. How many lines is that? Are there any lines left over, available to be "parallels"?)

(c) Prove that axiom B is satisfied.

37. Even in so miniature a geometry, we can do more than just play with points and lines. We can define circles, ellipses, hyperbolas, parabolas, and even tangents to circles! Here we'll just whet your appetite by introducing circles. A circle centered at (a, b) will consist of all of the points that satisfy an equation of the form $(x - a)^2 + (y - b)^2 = r$, where each of a, b, and r is one of 0, 1, or 2.

(a) Find the points on the circle $x^2 + y^2 = 0$.
(b) Find the points on the circle $x^2 + y^2 = 1$.
(c) Find the points on the circle $x^2 + y^2 = 2$.

38. [Adapted from Boltyansky (1992).] Take as the "points" of a geometry all points of the plane except a single point O. Take as "lines" all circles and straight lines that pass through the deleted point O.

(a) Given two distinct "points" A and B, show how to construct a "line" that passes through both of them. Is this the only line with this property?

(b) Given a "line" l and a "point" P not on l, show that there is a unique "line" through P that does not intersect l. In other words, the parallel postulate holds.

This geometry can be further furnished with angle measure, distance measure, "circles," triangles, and so forth, so that it satisfies all of the postulates of Euclidean geometry and hence is a model of those postulates.

Relativity and Length Contraction

39. For each of the following objects, calculate at what fraction of the speed of light it is traveling, the corresponding value of γ, and how much it is shortened by Lorentz-Fitzgerald contraction.

    (a) Automobile at 65 mph
    (b) Rifle bullet at 1 km/sec
    (e) Comet Hyakutake as it passed the earth at the end of March 1996, 200,000 mph

40. Repeat Exercise 39, but for

    (a) an object moving at 99% of the speed of light
    (b) subatomic particles in an accelerator at 99.94% of the speed of light

WRITING PROJECTS

For Projects 1 to 4, refer to the following (adapted from Davis [1968]): Folklore has it that everybody is (or should be) an enemy of their friends' enemies and a friend of their friends' friends, as well as a friend of their enemies' enemies and an enemy of their enemies' friends. If this is the case, what patterns of friendship are possible in a stable society? We investigate this question by converting the folklore into an axiom system and then seeing what conclusions can be drawn. We use capital letters to denote people, together with the symbols =, ♡, and ♯. Our interpretations of the symbols are

- $X = Y$ means X and Y are the same person.
- $X \heartsuit Y$ means that X is an immediate friend of Y.
- $X \sharp Y$ means that X is an immediate enemy of Y.

We also make some definitions. We say that X is *immediately involved with* Y if either $X \heartsuit Y$ or $X \sharp Y$. Also, we define a *chain of involvements* from X to Y as a sequence of people $Z_0, Z_1, \ldots, Z_n$ (for some n) such that $Z_0 = X$, $Z_n = Y$, and each Z_{i-1} is immediately involved with Z_i, for $i = 1, 2, \ldots, n$. The Z_i don't have to be distinct. In fact, we count $X \heartsuit X$ and $X \sharp X$ as chains of length one.

Our axioms are:

A For all X and Y, if $X \heartsuit Y$, then $Y \heartsuit X$.
B For all X and Y, if $X \sharp Y$, then $Y \sharp X$.
C Every pair of people is connected by a chain of involvements.

Axiom C asserts that no person is ever completely isolated from another. We call a chain of involvements *positive* if the number of immediate enmities in it is even, and *negative* if it is odd. We also say that X is a *(distant) friend* of Y if there is a positive chain of involvements from X to Y, and is a *(distant) enemy* if there is a negative chain. Finally, we say that X is *ambivalent* toward Y if X is both a friend and an enemy of Y.

1 ▸ Suppose that $X \heartsuit Z \sharp W \heartsuit Y$. Is X a friend of Y?

2 ▸ Show that if X is a friend of, an enemy of, or ambivalent toward Y, then Y is likewise related to X.

3 ▸ A *society* is a set of people, each pair of which is connected by a chain of involvements consisting of members of that set.

(a) Show that the set of all people is a society.
(b) Can one person alone be a society?

4 ▸ A society is *stable* if there exist no ambivalences in it: No pair of people are friends and enemies both.

(a) Show that in an unstable society, everyone is everyone's friend and enemy both. In other words, everybody is ambivalent toward everybody, including themselves.
(b) Show that a society is stable if and only if it divides into two sets of people so that everyone is a friend of just the people in their own set and an enemy of exactly those in the other. (One of the sets may be empty, in which case everyone is a friend of everyone and an enemy of none.)

5 ▸ Examine the axioms, terminology, and theorems of this axiomatic system. Discuss how well you think this theory fits human attitudes and behavior.