

Euclid

The story of axiomatic geometry begins with Euclid, the most famous mathematician in history. We know essentially nothing about Euclid's life, save that he was a Greek who lived and worked in Alexandria, Egypt, around 300 BCE. His best known work is the *Elements* [Euc02], a thirteen-volume treatise that organized and systematized essentially all of the knowledge of geometry and number theory that had been developed in the Western world up to that time.

It is believed that most of the mathematical results of the *Elements* were known well before Euclid's time. Euclid's principal achievement was not the discovery of new mathematical facts, but something much more profound: he was apparently the first mathematician to find a way to organize virtually all known mathematical knowledge into a single coherent, logical system, beginning with a list of definitions and a small number of assumptions (called *postulates*) and progressing logically to prove every other result from the postulates and the previously proved results. The *Elements* provided the Western world with a model of deductive mathematical reasoning whose essential features we still emulate today.

A brief remark is in order regarding the authorship of the *Elements*. Scholars of Greek mathematics are convinced that some of the text that has come down to us as the *Elements* was not in fact written by Euclid but instead was added by later authors. For some portions of the text, this conclusion is well founded—for example, there are passages that appear in earlier Greek manuscripts as marginal notes but that are part of the main text in later editions; it is reasonable to conclude that these passages were added by scholars after Euclid's time and were later incorporated into Euclid's text when the manuscript was recopied. For other passages, the authorship is less clear—some scholars even speculate that the definitions might have been among the later additions. We will probably never know exactly what Euclid's original version of the *Elements* looked like.

Since our purpose here is primarily to study the logical development of geometry and not its historical development, let us simply agree to use the name Euclid to refer to the writer or writers of the text that has been passed down to us as the *Elements* and leave it to the historians to explore the subtleties of multiple authorship.

Reading Euclid

Before going any further, you should take some time now to glance at Book I of the *Elements*, which contains most of Euclid’s elementary results about plane geometry. As we discuss each of the various parts of the text—definitions, postulates, common notions, and propositions—you should go back and read through that part carefully. Be sure to observe how the propositions build logically one upon another, each proof relying only on definitions, postulates, common notions, and previously proved propositions.

Here are some remarks about the various components of Book I.

Definitions

If you study Euclid’s definitions carefully, you will see that they can be divided into two rather different categories. Many of the definitions (including the first nine) are *descriptive definitions*, meaning that they are meant to convey to the reader an intuitive sense of what Euclid is talking about. For example, Euclid defines a *point* as “that which has no part,” a *line* as “breadthless length,” and a *straight line* as “a line which lies evenly with the points on itself.” (Here and throughout this book, our quotations from Euclid are taken from the well-known 1908 English translation of the *Elements* by T. L. Heath, based on the edition [Euc02] edited by Dana Densmore.) These descriptions serve to guide the reader’s thinking about these concepts but are not sufficiently precise to be used to justify steps in logical arguments because they typically define new terms in terms of other terms that have not been previously defined. For example, Euclid never explains what is meant by “breadthless length” or by “lies evenly with the points on itself”; the reader is expected to interpret these definitions in light of experience and common knowledge. Indeed, in all the books of the *Elements*, Euclid never refers to the first nine definitions, or to any other descriptive definitions, to justify steps in his proofs.

Contrasted with the descriptive definitions are the *logical definitions*. These are definitions that describe a precise mathematical condition that must be satisfied in order for an object to be considered an example of the defined term. The first logical definition in the *Elements* is Definition 10: “When a straight line standing on a straight line makes the adjacent angles equal to one another, each of the equal angles is *right*, and the straight line standing on the other is called a *perpendicular* to that on which it stands.” This describes angles in a particular type of geometric configuration (Fig. 1.1) and tells us that we are entitled to call an angle a *right angle* if and only if it occurs in a configuration of that type. (See Appendix E for a discussion about the use of “if and only if” in definitions.) Some other terms for which Euclid provides logical definitions are *circle*, *isosceles triangle*, and *parallel*.

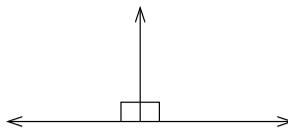


Fig. 1.1. Euclid’s definition of right angles.

Postulates

It is in the postulates that the great genius of Euclid's achievement becomes evident. Although mathematicians before Euclid had provided proofs of some isolated geometric facts (for example, the Pythagorean theorem was probably proved at least two hundred years before Euclid's time), it was apparently Euclid who first conceived the idea of arranging all the proofs in a strict logical sequence. Euclid realized that not every geometric fact can be proved, because every proof must rely on some prior geometric knowledge; thus any attempt to prove everything is doomed to circularity. He knew, therefore, that it was necessary to begin by accepting some facts without proof. He chose to begin by postulating five simple geometric statements:

- **Euclid's Postulate 1:** *To draw a straight line from any point to any point.*
- **Euclid's Postulate 2:** *To produce a finite straight line continuously in a straight line.*
- **Euclid's Postulate 3:** *To describe a circle with any center and distance.*
- **Euclid's Postulate 4:** *That all right angles are equal to one another.*
- **Euclid's Postulate 5:** *That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.*

The first three postulates are *constructions* and should be read as if they began with the words "It is possible." For example, Postulate 1 asserts that "[It is possible] to draw a straight line from any point to any point." (For Euclid, the term *straight line* could refer to a portion of a line with finite length—what we would call a *line segment*.) The first three postulates are generally understood as describing in abstract, idealized terms what we do concretely with the two classical geometric construction tools: a *straightedge* (a kind of idealized ruler that is unmarked but indefinitely extendible) and a *compass* (two arms connected by a hinge, with a sharp spike on the end of one arm and a drawing implement on the end of the other). With a straightedge, we can align the edge with two given points and draw a straight line segment connecting them (Postulate 1); and given a previously drawn straight line segment, we can align the straightedge with it and extend (or "produce") it in either direction to form a longer line segment (Postulate 2). With a compass, we can place the spike at any predetermined point in the plane, place the drawing tip at any other predetermined point, and draw a complete circle whose center is the first point and whose circumference passes through the second point. The statement of Postulate 3 does not precisely specify what Euclid meant by "any center and distance"; but the way he uses this postulate, for example in Propositions I.1 and I.2, makes it clear that it is applicable only when the center and one point on the circumference are already given. (In this book, we follow the traditional convention for referring to Euclid's propositions by number: "Proposition I.2" means Proposition 2 in Book I of the *Elements*.)

The last two postulates are different: instead of asserting that certain geometric configurations can be constructed, they describe relationships that must hold whenever a given geometric configuration exists. Postulate 4 is simple: it says that whenever two right angles have been constructed, those two angles are equal to each other. To interpret this, we must address Euclid's use of the word *equal*. In modern mathematical usage, " A equals B " just means the A and B are two different names for the same mathematical object (which could

be a number, an angle, a triangle, a polynomial, or whatever). But Euclid uses the word differently: when he says that two geometric objects are equal, he means essentially that they have the *same size*. In modern terminology, when Euclid says two angles are equal, we would say they have the same degree measure; when he says two lines (i.e., line segments) are equal, we would say they have the same length; and when he says two figures such as triangles or parallelograms are equal, we would say they have the same area. Thus Postulate 4 is actually asserting that all right angles are the same size.

It is important to understand why Postulate 4 is needed. Euclid's definition of a right angle applies only to an angle that appears in a certain configuration (one of the two adjacent angles formed when a straight line meets another straight line in such a way as to make equal adjacent angles); it does not allow us to conclude that a right angle appearing in one part of the plane bears any necessary relationship with right angles appearing elsewhere. Thus Postulate 4 can be thought of as an assertion of a certain type of "uniformity" in the plane: right angles have the same size wherever they appear.

Postulate 5 says, in more modern terms, that if one straight line crosses two other straight lines in such a way that the interior angles on one side have degree measures adding up to less than 180° ("less than two right angles"), then those two straight lines must meet on that same side of the first line (Fig. 1.2). Intuitively, it says that if two lines start out

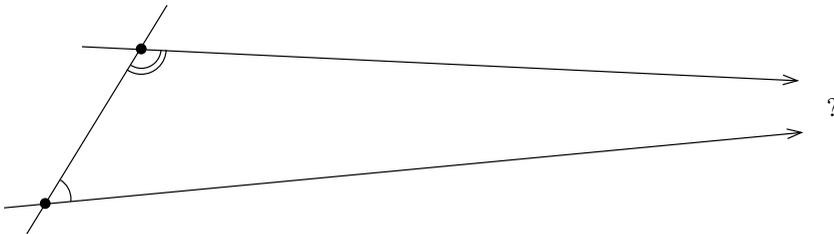


Fig. 1.2. Euclid's Postulate 5.

"pointing toward each other," they will eventually meet. Because it is used primarily to prove properties of parallel lines (for example, in Proposition I.29 to prove that parallel lines always make equal corresponding angles with a transversal), Euclid's fifth postulate is often called the "parallel postulate." We will have much more to say about it later in this chapter.

Common Notions

Following his five postulates, Euclid states five "common notions," which are also meant to be self-evident facts that are to be accepted without proof:

- **Common Notion 1:** *Things which are equal to the same thing are also equal to one another.*
- **Common Notion 2:** *If equals be added to equals, the wholes are equal.*
- **Common Notion 3:** *If equals be subtracted from equals, the remainders are equal.*
- **Common Notion 4:** *Things which coincide with one another are equal to one another.*
- **Common Notion 5:** *The whole is greater than the part.*

Whereas the five postulates express facts about geometric configurations, the common notions express facts about *magnitudes*. For Euclid, magnitudes are objects that can be compared, added, and subtracted, provided they are of the “same kind.” In Book I, the kinds of magnitudes that Euclid considers are (lengths of) line segments, (measures of) angles, and (areas of) triangles and quadrilaterals. For example, a line segment (which Euclid calls a “finite straight line”) can be equal to, greater than, or less than another line segment; two line segments can be added together to form a longer line segment; and a shorter line segment can be subtracted from a longer one.

It is interesting to observe that although Euclid compares, adds, and subtracts geometric magnitudes of the same kind, he never uses *numbers* to measure geometric magnitudes. This might strike one as curious, because human societies had been using numbers to measure things since prehistoric times. But there is a simple explanation for its omission from Euclid’s axiomatic treatment of geometry: to the ancient Greeks, *numbers* meant whole numbers, or at best ratios of whole numbers (what we now call *rational numbers*). However, the followers of Pythagoras had discovered long before Euclid that the relationship between the diagonal of a square and its side length cannot be expressed as a ratio of whole numbers. In modern terms, we would say that the ratio of the length of the diagonal of a square to its side length is equal to $\sqrt{2}$; but there is no rational number whose square is 2. Here is a proof that this is so.

Theorem 1.1 (Irrationality of $\sqrt{2}$). *There is no rational number whose square is 2.*

Proof. Assume the contrary: that is, assume that there are integers p and q with $q \neq 0$ such that $2 = (p/q)^2$. After canceling out common factors, we can assume that p/q is in lowest terms, meaning that p and q have no common prime factors. Multiplying the equation through by q^2 , we obtain

$$2q^2 = p^2. \quad (1.1)$$

Because p^2 is equal to the even number $2q^2$, it follows that p itself is even; thus there is some integer k such that $p = 2k$. Inserting this into (1.1) yields

$$2q^2 = (2k)^2 = 4k^2.$$

We can divide this equation through by 2 and obtain $q^2 = 2k^2$, which shows that q is also even. But this means that p and q have 2 as a common prime factor, contradicting our assumption that p/q is in lowest terms. Thus our original assumption must have been false. \square

This is one of the oldest examples of what we now call a *proof by contradiction* or *indirect proof*, in which we assume that a result is false and show that this assumption leads to a contradiction. A version of this argument appears in the *Elements* as Proposition VIII.8 (although it is a bit hard to recognize as such because of the archaic terminology Euclid used). For a more thorough discussion of proofs by contradiction, see Appendix F. For details of the properties of numbers that were used in the proof, see Appendix H.

This fact had the consequence that, for the Greeks, there was no “number” that could represent the length of the diagonal of a square whose sides have length 1. Thus it was not possible to assign a numeric length to every line segment.

Euclid’s way around this difficulty was simply to avoid using numbers to measure magnitudes. Instead, he only compares, adds, and subtracts magnitudes of the same kind.

(In later books, he also compares ratios of such magnitudes.) As mentioned above, it is clear from Euclid's use of the word "equal" that he always interprets it to mean "the same size"; any claim that two geometric figures are equal is ultimately justified by showing that one can be moved so that it coincides with the other or that the two objects can be decomposed into pieces that are equal for the same reason. His use of the phrases "greater than" and "less than" is always based on Common Notion 5: if one geometric object (such as a line segment or an angle) is part of another or is equal (in size) to part of another, then the first is less than the second.

Having laid out his definitions and assumptions, Euclid is now ready to start proving things.

Propositions

Euclid refers to every mathematical statement that he proves as a *proposition*. This is somewhat different from the usual practice in modern mathematical writing, where a result to be proved might be called a *theorem* (an important result, usually one that requires a relatively lengthy or difficult proof); a *proposition* (an interesting result that requires proof but is usually not important enough to be called a theorem); a *corollary* (an interesting result that follows from a previous theorem with little or no extra effort); or a *lemma* (a preliminary result that is not particularly interesting in its own right but is needed to prove another theorem or proposition).

Even though Euclid's results are all called propositions, the first thing one notices when looking through them is that, like the postulates, they are of two distinct types. Some propositions (such as I.1, I.2, and I.3) describe constructions of certain geometric configurations. (Traditionally, scholars of Euclid call these propositions *problems*. For clarity, we will call them *construction problems*.) These are usually stated in the infinitive ("to construct an equilateral triangle on a given finite straight line"), but like the first three postulates, they should be read as asserting the possibility of making the indicated constructions: "[It is possible] to construct an equilateral triangle on a given finite straight line."

Other propositions (traditionally called *theorems*) assert that certain relationships always hold in geometric configurations of a given type. Some examples are Propositions I.4 (the side-angle-side congruence theorem) and I.5 (the base angles of an isosceles triangle are equal). They do not assert the constructibility of anything. Instead, they apply only when a configuration of the given type has already been constructed, and they allow us to conclude that certain relationships always hold in that situation.

For both the construction problems and the theorems, Euclid's propositions and proofs follow a predictable pattern. Most propositions have six discernible parts. Here is how the parts were described by the Greek mathematician Proclus [Pro70]:

- (1) **Enunciation:** Stating in general form the construction problem to be solved or the theorem to be proved. Example from Proposition I.1: "On a given finite straight line to construct an equilateral triangle."
- (2) **Setting out:** Choosing a specific (but arbitrary) instance of the general situation and giving names to its constituent points and lines. Example: "Let AB be the given finite straight line."

- (3) **Specification:** Announcing what has to be constructed or proved in this specific instance. Example: “Thus it is required to construct an equilateral triangle on the straight line AB .”
- (4) **Construction:** Adding points, lines, and circles as needed. For construction problems, this is where the main construction algorithm is described. For theorems, this part, if present, describes any auxiliary objects that need to be added to the figure to complete the proof; if none are needed, it might be omitted.
- (5) **Proof:** Arguing logically that the given construction does indeed solve the given problem or that the given relationships do indeed hold.
- (6) **Conclusion:** Restating what has been proved.

A word about the conclusions of Euclid’s proofs is in order. Euclid and the classical mathematicians who followed him believed that a proof was not complete unless it ended with a precise statement of what had been shown to be true. For construction problems, this statement always ended with a phrase meaning “which was to be done” (translated into Latin as *quod erat faciendum*, or q.e.f.). For theorems, it ended with “which was to be demonstrated” (*quod erat demonstrandum*, or q.e.d.), which explains the origin of our traditional proof-ending abbreviation. In Heath’s translation of Proposition I.1, the conclusion reads “Therefore the triangle ABC is equilateral; and it has been constructed on the given finite straight line AB . Being what it was required to do.” Because this last step is so formulaic, after the first few propositions Heath abbreviates it: “Therefore etc. q.e.f.,” or “Therefore etc. q.e.d.”

We leave it to you to read Euclid’s propositions in detail, but it is worth focusing briefly on the first three because they tell us something important about Euclid’s conception of straightedge and compass constructions. Here are the statements of Euclid’s first three propositions:

Euclid’s Proposition I.1. *On a given finite straight line to construct an equilateral triangle.*

Euclid’s Proposition I.2. *To place a straight line equal to a given straight line with one end at a given point.*

Euclid’s Proposition I.3. *Given two unequal straight lines, to cut off from the greater a straight line equal to the less.*

One might well wonder why Euclid chose to start where he did. The construction of an equilateral triangle is undoubtedly useful, but is it really more useful than other fundamental constructions such as bisecting an angle, bisecting a line segment, or constructing a perpendicular? The second proposition is even more perplexing: all it allows us to do is to construct a copy of a given line segment with one end at a certain predetermined point, but we have no control over which direction the line segment points. Why should this be of any use whatsoever?

The mystery is solved by the third proposition. If you look closely at the way Postulate 3 is used in the first two propositions, it becomes clear that Euclid has a very specific interpretation in mind when he writes about “describing a circle with any center and distance.” In Proposition I.1, he describes the circle with center A and distance AB and the circle with center B and distance BA ; and in Proposition I.2, he describes circles with center B and distance BC and with center D and distance DG . In every case, the center is a point that

and failed to come up with proofs that the fifth postulate follows logically from the other four, or at least to replace it with a more truly self-evident postulate. This quest, in fact, has motivated much of the development of geometry since Euclid.

The earliest attempt to prove the fifth postulate that has survived to modern times was by Proclus (412–485 CE), a Greek philosopher and mathematician who lived in Asia Minor during the time of the early Byzantine empire and wrote an important commentary on Euclid’s *Elements* [Pro70]. (This commentary, by the way, contains most of the scant biographical information we have about Euclid, and even this must be considered essentially as legendary because it was written at least 700 years after the time of Euclid.) In this commentary, Proclus opined that Postulate 5 did not have the self-evident nature of a postulate and thus should be proved, and then he proceeded to offer a proof of it. Unfortunately, like so many later attempts, Proclus’s proof was based on an unstated and unproved assumption. Although Euclid *defined* parallel lines to be lines in the same plane that do not meet, no matter how far they are extended, Proclus tacitly assumed also that parallel lines are everywhere *equidistant*, meaning the same distance apart (see Fig. 1.3). We will see in Chapter 17 that this assumption is actually equivalent to assuming Euclid’s fifth postulate.

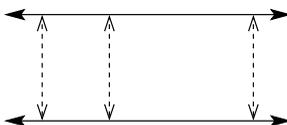


Fig. 1.3. Proclus’s assumption.

After the fall of the Roman Empire, the study of geometry moved for the most part to the Islamic world. Although the original Greek text of Euclid’s *Elements* was lost until the Renaissance, translations into Arabic were widely studied throughout the Islamic empire and eventually made their way back to Europe to be translated into Latin and other languages.

During the years 1000–1300, several important Islamic mathematicians took up the study of the fifth postulate. Most notable among them was the Persian scholar and poet Omar Khayyam (1048–1123), who criticized previous attempts to prove the fifth postulate and then offered a proof of his own. His proof was incorrect because, like that of Proclus, it relied on the unproved assumption that parallel lines are everywhere equidistant.

With the advent of the Renaissance, Western Europeans again began to tackle the problem of the fifth postulate. One of the most important attempts was made by the Italian mathematician Giovanni Saccheri (1667–1733). Saccheri set out to prove the fifth postulate by assuming that it was false and showing that this assumption led to a contradiction. His arguments were carefully constructed and quite rigorous for their day. In the process, he proved a great many strange and counterintuitive theorems that follow from the assumption that the fifth postulate is false, such as that rectangles cannot exist and that the interior angle measures of triangles always add up to less than 180° . In the end, though, he could not find a contradiction that measured up to the standards of rigor he had set for himself. Instead, he punted: having shown that his assumption implied that there must exist parallel lines that approach closer and closer to each other but never meet, he claimed that this result is “repugnant to the nature of the straight line” and therefore his original assumption must have been false.

Saccheri is remembered today, not for his failed attempt to prove Euclid's fifth postulate, but because in attempting to do so he managed to prove a great many results that we now recognize as theorems in a mysterious new system of geometry that we now call *non-Euclidean geometry*. Because of the preconceptions built into the cultural context within which he worked, he could only see them as steps along the way to his hoped-for contradiction, which never came.

The next participant in our story played a minor role, but a memorable one. In 1795, the Scottish mathematician John Playfair (1748–1819) published an edition of the first six books of Euclid's *Elements* [Pl95], which he had edited to correct some of what were then perceived as flaws in the original. One of Playfair's modifications was to replace Euclid's fifth postulate with the following alternative postulate.

Playfair's Postulate. *Two straight lines cannot be drawn through the same point, parallel to the same straight line, without coinciding with one another.*

In other words, given a line and a point not on that line, there can be at most one line through the given point and parallel to the given line. Playfair showed that this alternative postulate leads to the same conclusions as Euclid's fifth postulate. This postulate has a notable advantage over Euclid's fifth postulate, because it focuses attention on the uniqueness of parallel lines, which (as later generations were to learn) is the crux of the issue. Most modern treatments of Euclidean geometry incorporate some version of Playfair's postulate instead of the fifth postulate originally stated by Euclid.

The Discovery of Non-Euclidean Geometry

The next event in the history of geometry was the most profound development in mathematics since the time of Euclid. In the 1820s, a revolutionary idea occurred independently and more or less simultaneously to three different mathematicians: perhaps the reason the fifth postulate had turned out to be so hard to prove was that there is a completely consistent theory of geometry in which Euclid's first four postulates are true but the fifth postulate is *false*. If this speculation turned out to be justified, it would mean that proving the fifth postulate from the other four would be a logical impossibility.

In 1829, the Russian mathematician Nikolai Lobachevsky (1792–1856) published a paper laying out the foundations of what we now call *non-Euclidean geometry*, in which the fifth postulate is assumed to be false, and proving a good number of theorems about it. Meanwhile in Hungary, János Bolyai (1802–1860), the young son of an eminent Hungarian mathematician, spent the years 1820–1823 writing a manuscript that accomplished much the same thing; his paper was eventually published in 1832 as an appendix to a textbook written by his father. When the great German mathematician Carl Friedrich Gauss (1777–1855)—a friend of Bolyai's father—read Bolyai's paper, he remarked that it duplicated investigations of his own that he had carried out privately but never published. Although Bolyai and Lobachevsky deservedly received the credit for having invented non-Euclidean geometry based on their published works, in view of the creativity and depth of Gauss's other contributions to mathematics, there is no reason to doubt that Gauss had indeed had the same insight as Lobachevsky and Bolyai.

In a sense, the principal contribution of these mathematicians was more a change of attitude than anything else: while Omar Khayyam, Giovanni Saccheri, and others had also

proved theorems of non-Euclidean geometry, it was Lobachevsky and Bolyai (and presumably also Gauss) who first recognized them as such. However, even after this groundbreaking work, there was still no *proof* that non-Euclidean geometry was consistent (i.e., that it could never lead to a contradiction). The *coup de grâce* for attempts to prove the fifth postulate was provided in 1868 by another Italian mathematician, Eugenio Beltrami (1835–1900), who proved for the first time that non-Euclidean geometry was just as consistent as Euclidean geometry. Thus the ancient question of whether Euclid’s fifth postulate followed from the other four was finally settled.

The versions of non-Euclidean geometry studied by Lobachevsky, Bolyai, Gauss, and Beltrami were all essentially equivalent to each other. This geometry is now called *hyperbolic geometry*. Its most salient feature is that Playfair’s postulate is false: in hyperbolic geometry it is always possible for two or more distinct straight lines to be drawn through the same point, both parallel to a given straight line. As a consequence, many aspects of Euclid’s theory of parallel lines (such as the result in Proposition I.29 about the equality of corresponding angles made by a transversal to two parallel lines) are not valid in hyperbolic geometry. In fact, as we will see in Chapter 19, the phenomenon of parallel lines approaching each other asymptotically but never meeting—which Saccheri declared to be “repugnant to the nature of the straight line”—does indeed occur in hyperbolic geometry.

One might also wonder if the Euclidean theory of parallel lines could fail in the opposite way: instead of having two or more parallels through the same point, might it be possible to construct a consistent theory in which there are *no* parallels to a given line through a given point? It is easy to imagine a type of geometry in which there are no parallel lines: the geometry of a sphere. If you move as straight as possible on the surface of a sphere, you will follow a path known as a *great circle*—a circle whose center coincides with the center of the sphere. It can be visualized as the place where the sphere intersects a plane that passes through the center of the sphere. If we reinterpret the term “line” to mean a great circle on the sphere, then indeed there are no parallel “lines,” because any two great circles must intersect each other. But this does not seem to have much relevance for Euclid’s geometry, because line segments cannot be extended arbitrarily far—in spherical geometry, no line can be longer than the circumference of the sphere. This would seem to contradict Postulate 2, which had always been interpreted to mean that a line segment can be extended arbitrarily far in both directions.

However, after the discovery of hyperbolic geometry, another German mathematician, Bernhard Riemann (1826–1866), realized that Euclid’s second postulate could be reinterpreted in such a way that it does hold on the surface of a sphere. Basically, he argued that Euclid’s second postulate only requires that any line segment can be extended to a longer one in both directions, but it does not specifically say that we can extend it to any length we wish. With this reinterpretation, spherical geometry can be seen to be a consistent geometry in which no lines are parallel to each other. Of course, a number of Euclid’s proofs break down in this situation, because many of the implicit geometric assumptions he used in his proofs do not hold on the sphere; see our discussion of Euclid’s Proposition I.16 below for an example. This alternative form of non-Euclidean geometry is sometimes called *elliptic geometry*. (Because of its association with Riemann, it is sometimes erroneously referred to as *Riemannian geometry*, but this term is now universally used to refer to an entirely different type of geometry, which is a branch of differential geometry.)

Perhaps the most convincing confirmation that Euclid's is not the only possible consistent theory of geometry came from Einstein's general theory of relativity around the turn of the twentieth century. If we are to believe, like Euclid, that the postulates reflect self-evident truths about the geometry of the world we live in, then Euclid's statements about "straight lines" should translate into true statements about the behavior of light rays in the real world. (After all, we commonly judge the "straightness" of something by sighting along it, so what physical phenomenon could possibly qualify as a better model of "straight lines" than light rays?) Thus the closest thing in the physical world to a geometric triangle would be a three-sided figure whose sides are formed by the paths of light rays.

Yet Einstein's theory tells us that in the presence of gravitational fields, space itself is "warped," and this affects the paths along which light rays travel. One of the most dramatic confirmations of Einstein's theory comes from the phenomenon known as *gravitational lensing*: this occurs when we observe a distant object but there is a massive galaxy cluster directly between us and the object. Einstein's theory predicts that the light rays from the distant object should be able to follow two (or more) different paths to reach our eyes because of the distortion of space around the galaxy cluster.

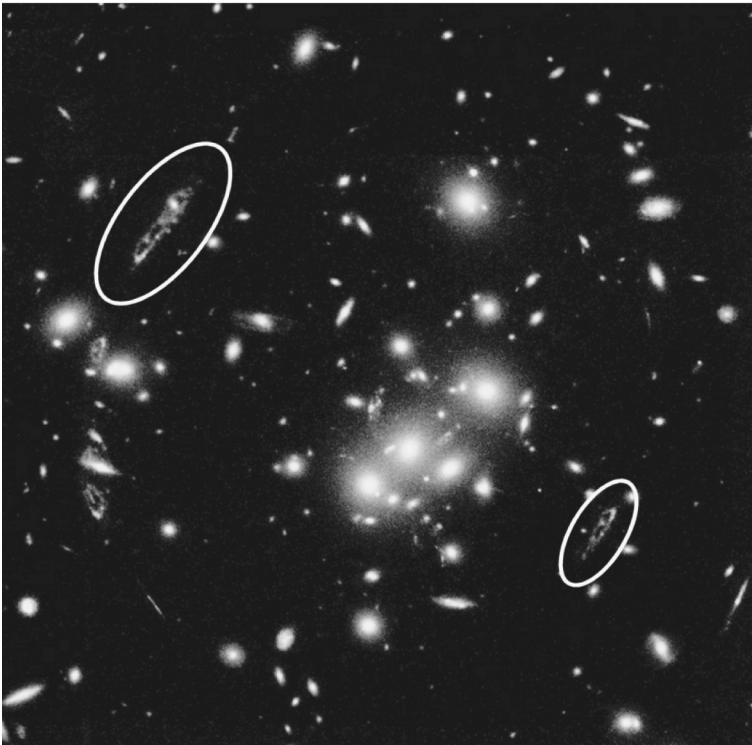


Fig. 1.4. A gravitational lens (photograph by W. N. Colley, E. Turner, J. A. Tyson, and NASA).

This phenomenon has indeed been observed; Fig. 1.4 shows a photograph taken by the Hubble Space Telescope, in which a distant loop-shaped galaxy (circled) appears twice in the same photographic image because its light rays have traveled around both sides of the large galaxy cluster in the middle of the photo. Fig. 1.5 shows a schematic view of the same

situation. The light coming from a certain point A in the middle of the loop-shaped galaxy follows two paths to our eyes and along the way makes two different dots (B and C) on the photographic plate. As a result, the three points A , B , and C form a triangle whose interior angle measures add up to a number slightly greater than 180° . (Although it doesn't look like a triangle in this diagram, remember that the edges are paths of light rays. What could be straighter than that?) Yet Euclidean geometry predicts that every triangle has interior angle measures that add up to *exactly* 180° . We can see why Euclid's arguments fail in this situation by examining the figure: in this case there are two distinct line segments connecting the point A to the observer's eye, which contradicts Euclid's intended meaning of his first postulate. We have no choice but to conclude that the geometry of the physical world we live in does not exactly follow Euclid's rules.

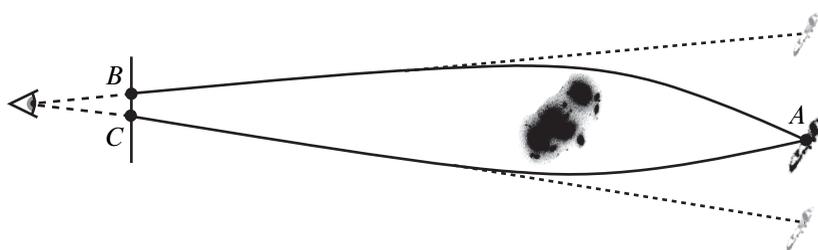


Fig. 1.5. A triangle whose angle sum is greater than 180° .

Gaps in Euclid's Arguments

As a result of the non-Euclidean breakthroughs of Lobachevsky, Bolyai, and Gauss, mathematicians were forced to undertake a far-reaching reexamination of the very foundations of their subject. Euclid and everyone who followed him had regarded postulates as self-evident truths about the real world, from which reliable conclusions could be drawn. But once it was discovered that two or three conflicting systems of postulates worked equally well as logical foundations for geometry, mathematicians had to face an uncomfortable question: what exactly are we doing when we accept some postulates and use them to prove theorems? It became clear that the system of postulates one uses is in some sense an arbitrary choice; once the postulates have been chosen, as long as they don't lead to any contradictions, one can proceed to prove whatever theorems follow from them.

Thus was born the notion of a mathematical theory as an *axiomatic system*—a sequence of theorems based on a particular set of assumptions called *postulates* or *axioms* (these two words are used synonymously in modern mathematics). We will give a more precise definition of axiomatic systems in the next chapter.

Of course, the axioms we choose are not *completely* arbitrary, because the only axiomatic systems that are worth studying are those that describe something useful or interesting—an aspect of the physical world, or a class of mathematical structures that have proved useful in other contexts, for example. But from a strictly logical point of view, we may adopt any consistent system of axioms that we like, and the resulting theorems will constitute a valid mathematical theory.

The catch is that one must scrupulously ensure that the proofs of the theorems do not use *anything* other than what has been assumed in the postulates. If the axioms represent arbitrary assumptions instead of self-evident facts about the real world, then nothing except the axioms is relevant to proofs within the system. Reasoning based on intuition about the behavior of straight lines or properties that are evident from diagrams or common experience in the real world will no longer be justifiable within the axiomatic system.

Looking back at Euclid with these newfound insights, mathematicians realized that Euclid had used many properties of lines and circles that were not strictly justified by his postulates. Let us examine a few of those properties, as a way of motivating the more careful axiomatic system that we will develop later in the book. We will discuss some of the most problematic of Euclid's proofs in the order in which they occur in Book I. As always, we refer to the edition [Euc02].

While reading these analyses of Euclid's arguments, you should bear in mind that we are judging the incompleteness of these proofs based on criteria that would have been utterly irrelevant in Euclid's time. For the ancient Greeks, geometric proofs were meant to be convincing arguments about the geometry of the physical world, so basing geometric conclusions on facts that were obvious from diagrams would never have struck them as an invalid form of reasoning. Thus these observations should not be seen as criticisms of Euclid; rather, they are meant to help point the way toward the development of a new axiomatic system that lives up to our modern (post-Euclidean) conception of rigor.

Euclid's Proposition I.1. *On a given finite straight line to construct an equilateral triangle.*

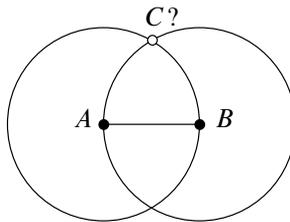


Fig. 1.6. Euclid's proof of Proposition I.1.

Analysis. In Euclid's proof of this, his very first proposition, he draws two circles, one centered at each endpoint of the given line segment AB (see Fig. 1.6). (In geometric diagrams in this book, we will typically draw selected points as small black dots to emphasize their locations; this is merely a convenience and is not meant to suggest that points take up any area in the plane.) He then proceeds to mention "the point C , in which the circles cut one another." It seems obvious from the diagram that there is a point where the circles intersect, but which of Euclid's postulates justifies the fact that such a point always exists? Notice that Postulate 5 asserts the existence of point where two *lines* intersect under certain circumstances; but nowhere does Euclid give us any justification for asserting the existence of a point where two *circles* intersect.

Euclid's Proposition I.3. *Given two unequal straight lines, to cut off from the greater a straight line equal to the less.*

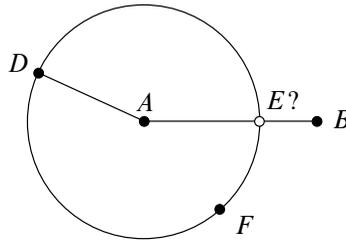


Fig. 1.7. Euclid's proof of Proposition I.3.

Analysis. In his third proof, Euclid implicitly uses another unjustified property of circles, although this one is a little more subtle. Starting with a line segment AD that he has just constructed, which shares an endpoint with a longer line segment AB , he draws a circle DEF with center A and passing through D (justified by Postulate 3). Although he does not say so explicitly, it is evident from his drawing that he means for E to be a point that is simultaneously on the circle DEF and also on the line segment AB . But once again, there is nothing in his list of postulates (or in the two previously proved propositions) that justifies the claim that a circle must intersect a line. (The same unjustified step also occurs twice in the proof of Proposition I.2, but it is a little easier to see in Proposition I.3.)

Euclid's Proposition I.4. *If two triangles have the two sides equal to two sides respectively, and have the angles contained by the equal straight lines equal, they will also have the base equal to the base, the triangle will be equal to the triangle, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend.*

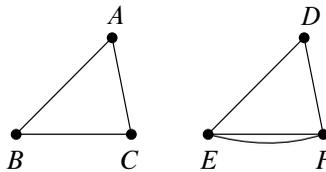


Fig. 1.8. Euclid's proof of Proposition I.4.

Analysis. This is Euclid's proof of the well-known *side-angle-side congruence theorem* (SAS). He begins with two triangles, ABC and DEF , such that $AB = DE$, $AC = DF$, and angle BAC is equal to angle EDF . (For the time being, we are adopting Euclid's convention that "equal" means "the same size.") He then says that triangle ABC should be "applied to triangle DEF ." (Some revised translations use "superposed upon" or "superimposed upon" in place of "applied to.") The idea is that we should imagine triangle ABC being moved over on top of triangle DEF in such a way that A lands on D and the segment AB points in the same direction as DE , so that the moved-over copy of ABC occupies the same position in the plane as DEF . (Although Euclid does not explicitly mention it, he also evidently intends for C to be placed on the same side of the line DE as F , to ensure that the moved-over copy of ABC will coincide with DEF instead of being a mirror image of it.) This technique has become known as the *method of superposition*.

This is an intuitively appealing argument, because we have all had the experience of moving cutouts of geometric figures around to make them coincide. However, there is nothing in Euclid's postulates that justifies the claim that a geometric figure can be moved, much less that its geometric properties such as side lengths and angle measures will remain unchanged after the move. Of course, Propositions I.2 and I.3 describe ways of constructing "copies" of line segments at other positions in the plane, but they say nothing about copying angles or triangles. (In fact, he does prove later, in Proposition I.23, that it is possible to construct a copy of an *angle* at a different location; but that proof depends on Proposition I.4!)

This is one of the most serious gaps in Euclid's proofs. In fact, many scholars have inferred that Euclid himself was uncomfortable with the method of superposition, because he used it in only three proofs in the entire thirteen books of the *Elements* (Propositions I.4, I.8, and III.23), despite the fact that he could have simplified many other proofs by using it.

There is another important gap in Euclid's reasoning in this proof: having argued that triangle ABC can be moved so that A coincides with D , B coincides with E , and C coincides with F , he then concludes that the line segments BC and EF will also coincide and hence be equal (in size). Now, Postulate 1 says that it is possible to construct a straight line (segment) from any point to any other point, but it does not say that it is possible to construct *only one* such line segment. Thus the postulates provide no justification for concluding that the segments BC and EF will necessarily coincide, even though they have the same endpoints. Euclid evidently meant the reader to understand that there is a *unique* line segment from one point to another point. In a modern axiomatic system, this would have to be stated explicitly.

Euclid's Proposition I.10. *To bisect a given finite straight line.*

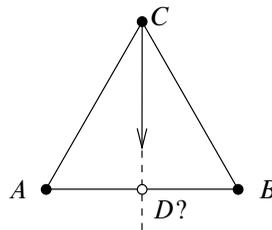


Fig. 1.9. Euclid's proof of Proposition I.10.

Analysis. In the proof of this proposition, Euclid uses another subtle property of intersections that is not justified by the postulates. Given a line segment AB , he constructs an equilateral triangle ABC with AB as one of its sides (which is justified by Proposition I.1) and then constructs the bisector of angle ACB (justified by Proposition I.9, which he has just proved). So far, so good. But his diagram shows the angle bisector intersecting the segment AB at a point D , and he proceeds to prove that AB is bisected (or cut in half) at this very point D . Once again, there is nothing in the postulates that justifies Euclid's assertion that there must be such an intersection point.

Euclid's Proposition I.12. *To a given infinite straight line, from a given point which is not on it, to draw a perpendicular straight line.*

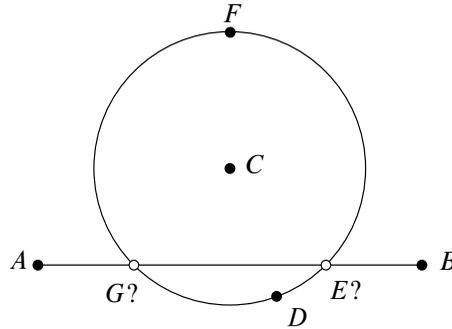


Fig. 1.10. Euclid's proof of Proposition I.12.

Analysis. In this proof, Euclid starts with a line AB and a point C not on that line. He then says, "Let a point D be taken at random on the other side of the straight line AB , and with center C and distance CD let the circle EFG be described." He is stipulating that the circle should be drawn with D on its circumference, which is exactly what Postulate 3 allows one to do. However, he is also implicitly assuming that such a circle will intersect AB in two points, which he calls E and G . Obviously it is the fact that C and D are on opposite sides of AB that is supposed to guarantee the existence of the intersection points; but which of his postulates or previous propositions justifies this conclusion? For that matter, what is "on the other side" supposed to mean? Euclid's definitions and postulates do not mention "sides" of lines at all, but he regularly refers to them in his proofs. It is clear from the diagrams what he means, but it is not justified by the postulates.

Euclid's Proposition I.16. *In any triangle, if one of the sides be produced, the exterior angle is greater than either of the interior and opposite angles.*

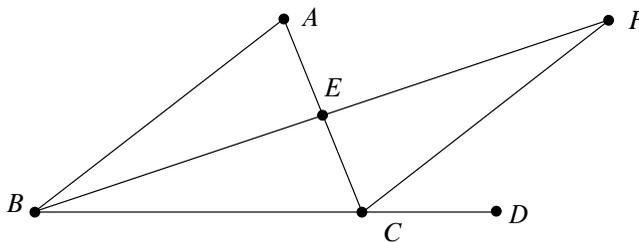


Fig. 1.11. Euclid's proof of Proposition I.16.

Analysis. Nowadays, this result is called the *exterior angle inequality*. Its proof is one of the most subtle and clever in the *Elements*. It is worth reading it over once or twice to absorb the full impact.

It is not easy to see where the gaps are, but there are at least two. After constructing the point E that bisects AC (using Proposition I.10), Euclid then extends BE past E and

uses Proposition I.3 to choose a point F on that line such that EF is the same length as BE . Here is where the first problem arises: although Postulate 3 guarantees that a line segment can be extended to form a longer line segment containing the original one, it does not explicitly say that we can make the extended line segment as long as we wish. As we mentioned above, if we were working on the surface of a sphere, this might not be possible because great circles have a built-in maximum length.

The second problem arises toward the end of the proof, when Euclid claims that angle ECD is greater than angle ECF . This is supposed to be justified by Common Notion 5 (the whole is greater than the part). However, in order to claim that angle ECF is “part of” angle ECD , we need to know that F lies in the interior of angle ECD . This seems evident from the diagram, but once again, there is nothing in the axioms or previous propositions that justifies the claim. To see how this could fail, consider once again the surface of a sphere. In Fig. 1.12, we have illustrated an analogous configuration, with A at the north pole and B and C both on the equator. If B and C are far enough apart, it is entirely possible for the point F to end up south of the equator, in which case it is no longer in the interior of angle ECD . (Fig. 1.13 illustrates the same configuration after it has been “unwrapped” onto a plane.)

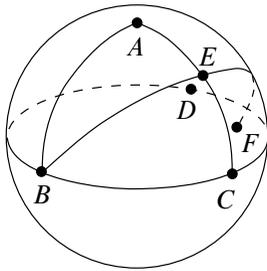


Fig. 1.12. Euclid’s proof fails on a sphere.

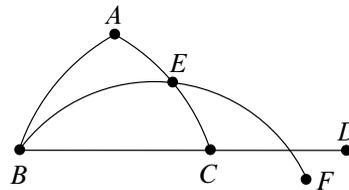


Fig. 1.13. The same diagram “unwrapped.”

Some of these objections to Euclid’s arguments might seem to be of little practical consequence, because, after all, nobody questions the truth of the theorems he proved. However, if one makes a practice of relying on relationships that seem obvious in diagrams, it is possible to go wildly astray. We end this section by presenting a famous fallacious “proof” of a false “theorem,” which vividly illustrates the danger.

The argument below is every bit as rigorous as Euclid’s proofs, with each step justified by Euclid’s postulates, common notions, or propositions; and yet the theorem being proved is one that everybody knows to be false. This proof was first published in 1892 in a recreational mathematics book by W. W. Rouse Ball [Bal87, p. 48]. Exercise 1D asks you to identify the incorrect step(s) in the proof.

Fake Theorem. *Every triangle has at least two equal sides.*

Fake Proof. Let ABC be any triangle, and let AD be the bisector of angle A (Proposition I.9). We consider several cases.

Suppose first that when AD is extended (Postulate 2), it meets BC perpendicularly. Let O be the point where these segments meet (Fig. 1.14(a)). Then angles AOB and AOC are both right angles by definition of “perpendicular.” Thus the triangles AOB and

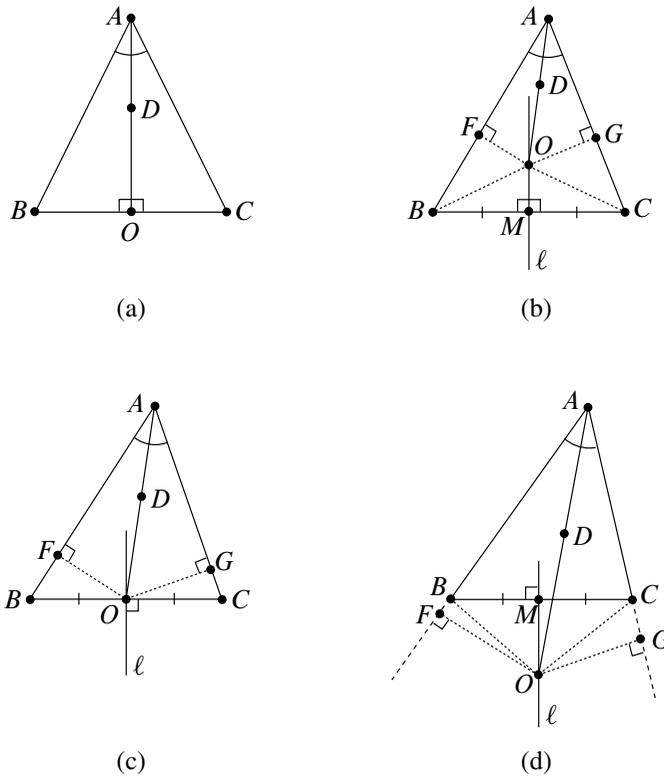


Fig. 1.14. The “proof” that every triangle has two equal sides.

AOC have two pairs of equal angles and share the common side AO , so it follows from Proposition I.26 that the sides AB and AC are equal.

In all of the remaining cases, we assume that the extension of AD is not perpendicular to BC . Let BC be bisected at M (Proposition I.10), let ℓ be the line perpendicular to BC at M (Proposition I.11), and let AD be extended if necessary (Postulate 2) so that it meets ℓ at O . There are now three possible cases, depending on the location of O .

CASE 1: O lies inside triangle ABC (Fig. 1.14(b)). Draw BO and CO (Postulate 1). Note that the triangles BMO and CMO have two pairs of equal corresponding sides (MO is common and $BM = CM$), and the included angles BMO and CMO are both right, so the remaining sides BO and CO are also equal by Proposition I.4. Now draw OF perpendicular to AB and OG perpendicular to AC (Proposition I.12). Then triangles AFO and AGO have two pairs of equal corresponding angles and share the side AO , so Proposition I.26 implies that their remaining pairs of corresponding sides are equal: $AF = AG$ and $FO = GO$. Now we can conclude that BFO and CGO are both right triangles in which the hypotenuses BO and CO are equal and the legs FO and GO are also equal. Therefore, the Pythagorean theorem (Proposition I.47) together with Common

Notion 3 implies that the squares on the remaining legs FB and GC are equal, and thus the legs themselves are also equal. Thus we have shown $AF = AG$ and $FB = GC$, so by Common Notion 2, it follows that $AB = AC$.

CASE 2: O lies on BC (Fig. 1.14(c)). Then O must be the point where BC is bisected, because that is where ℓ meets BC . In this case, we argue exactly as in Case 1, except that we can skip the first step involving triangles BMO and CMO , because we already know that $BO = OC$ (because BC is bisected at O). The rest of the proof proceeds exactly as before to yield the conclusion that $AB = AC$.

CASE 3: O lies outside triangle ABC (Fig. 1.14(d)). Again, the proof proceeds exactly as in Case 1, except now there are two changes: first, before drawing OF and OG , we need to extend AB beyond B , extend AC beyond C (Postulate 2), and draw OF and OG perpendicular to the extended line segments (Proposition I.12). Second, in the very last step, having shown that $AF = AG$ and $FB = GC$, we now use Common Notion 3 instead of Common Notion 2 to conclude that $AB = AC$. \square

Modern Axiom Systems

We have seen that the discovery of non-Euclidean geometry made it necessary to rethink the foundations of geometry, even Euclidean geometry. In 1899, these efforts culminated in the development by the German mathematician David Hilbert (1862–1943) of the first set of postulates for Euclidean geometry that were sufficient (according to modern standards of rigor) to prove all of the propositions in the *Elements*. (One version of Hilbert’s axioms is reproduced in Appendix A.) Following the tradition established by Euclid, Hilbert did not refer to numbers or measurements in his axiom system. In fact, he did not even refer to comparisons such as “greater than” or “less than”; instead, he introduced new relationships called *congruent* and *between* and added a number of axioms that specify their properties. For example, two line segments are to be thought of as congruent to each other if they have the same length (Euclid would say they are “equal”); and a point B is to be thought of as between A and C if B is an interior point of the segment AC (Euclid would say “ AB is part of AC ”). But these intuitive ideas were only the motivations for the choice of terms; the only facts about these terms that could legitimately be used in proofs were the facts stated explicitly in the axioms, such as Hilbert’s Axiom II.3: *Of any three points on a line there exists no more than one that lies between the other two.*

Although Hilbert’s axioms effectively filled in all of the unstated assumptions in Euclid’s arguments, they had a distinct disadvantage in comparison with Euclid’s postulates: Hilbert’s list of axioms was long and complicated and seemed to have lost the elegant simplicity of Euclid’s short list of assumptions. One reason for this complexity was the necessity of spelling out all the properties of betweenness and congruence that were needed to justify all of Euclid’s assertions regarding comparisons of magnitudes.

In 1932, the American mathematician George D. Birkhoff published a completely different set of axioms for plane geometry using real numbers to measure sizes of line segments and angles. The theoretical foundations of the real numbers had by then been solidly established, and Birkhoff reasoned that since numerical measurements are used ubiquitously in practical applications of geometry (as embodied in the ruler and protractor), there was no longer any good reason to exclude them from the axiomatic foundations of

geometry. Using these ideas, he was able to replace Hilbert's long list by only four axioms (see Appendix B).

Once Birkhoff's suggestion started to sink in, high-school text writers soon came around. Beginning with a textbook coauthored by Birkhoff himself [BB41], many high-school geometry texts were published in the U.S. that adopted axiom systems based more or less on Birkhoff's axioms. In the 1960s, the School Mathematics Study Group (MSG), a committee sponsored by the U.S. National Science Foundation, developed an influential system of axioms for high-school courses that used the real numbers in the way that Birkhoff had proposed (see Appendix C). The use of numbers for measuring lengths and angles was embodied in two axioms that the MSG authors called the *ruler postulate* and the *angle measurement postulate*. In one way or another, the MSG axioms form the basis for the axiomatic systems used in most high-school geometry texts today. The axioms that will be used in this book are inspired by the MSG axioms, although they have been modified in various ways: some of the redundant axioms have been eliminated, and some of the others have been rephrased to more closely capture our intuitions about plane geometry.

This concludes our brief survey of the historical events leading to the development of the modern axiomatic method. For a detailed and engaging account of the history of geometry from Euclid to the twentieth century, the book [Gre08] is highly recommended.

Exercises

- 1A. Read all of the definitions in Book I of Euclid's *Elements*, and identify which ones are descriptive and which are logical.
- 1B. Copy Euclid's proofs of Propositions I.6 and I.10, and identify each of the standard six parts: enunciation, setting out, specification, construction, proof, and conclusion.
- 1C. Choose several of the propositions in Book I of the *Elements*, and rewrite the statement and proof of each in more modern, idiomatic English. (You are not being asked to change the proofs or to fill in any of the gaps; all you need to do is rephrase Euclid's proofs to make them more understandable to modern readers.) When you do your rewriting, consider the following:
 - Be sure to include diagrams, and consider adding additional diagrams if they would help the reader follow the arguments.
 - Many terms are used by Euclid without explanation, so make sure you know what he means by them. The following terms, for example, are used frequently by Euclid but seldom in modern mathematical writing, so once you understand what they mean, you should consider replacing them by more commonly understood terms:
 - a *finite straight line*,
 - to *produce* a finite straight line,
 - to *describe* a circle,
 - to *apply* or *superpose* a figure onto another,
 - the *base* and *sides* of an arbitrary triangle,
 - one angle or side is *much greater* than another,
 - a straight line *standing on* a straight line,
 - an angle *subtended* by a side of a triangle.

- In addition, the following terms that Euclid uses without explanation are also used by modern writers, so you don't necessarily need to change them; but make sure that you know what they mean and that the meanings will be clear to your readers:
 - the *base* of an isosceles triangle,
 - to *bisect* a line segment or an angle,
 - *vertical angles*,
 - *adjacent angles*,
 - *exterior angles*,
 - *interior angles*.
 - Euclid sometimes writes “I say that [something is true],” which is a phrase you will seldom find in modern mathematical writing. When you see this phrase in Euclid, think about how it fits into the logic of his proof. Is he saying this is a statement that follows from what he has already proved? Or a statement that he thinks is obvious and does not need proof? Or a statement that he claims to be true but has not proved yet? Or something else? How might a modern mathematician express this?
 - Finally, after you have rewritten each proof, write a short discussion of the main features of the proof, and try to answer these questions: Why did Euclid construct the proof as he did? Were there any steps that seemed superfluous to you? Were there any steps or justifications that he left out? Why did this proposition appear at this particular place in the *Elements*? What would have been the consequences of trying to prove it earlier or later?
- 1D. Identify the fallacy that invalidates the proof of the “fake theorem” that says every triangle has two equal sides, and justify your analysis by carefully drawing an example of a nonisosceles triangle in which that step is actually false. [Hint: The problem has to do with drawing conclusions from the diagrams about locations of points. It's not enough just to find a step that is not adequately justified by the axioms; you must find a step that is actually false.]
- 1E. Find a modern secondary-school geometry textbook that includes some treatment of axioms and proofs, and do the following:
- (a) Read the first few chapters, including at least the chapter that introduces triangle congruence criteria (SAS, ASA, AAS).
 - (b) Do the homework exercises in the chapter that introduces triangle congruence criteria.
 - (c) Explain whether the axioms used in the book fill in some or all of the gaps in Euclid's reasoning discussed in this chapter.