4 A Critique of Euclid

It may seem anachronistic to return to a discussion of Euclid at this point. It is an old saw in the rules of teaching that one should never teach mistakes, should not use errors or false statements as illustrations. But, while it is certainly not wise to reinforce errors by repeating them, let us counter one old saying with another: it is always possible to learn from past mistakes.

We are not returning to Euclid for the sole purpose of pointing out his mistakes. Rather, we hope to prepare the reader for what is to come in the following chapters.

The emphasis today on abstraction in mathematics sometimes causes mathematicians to overlook that simple understanding called intuition. When this happens, a brief look at the source of an idea may be helpful. We hope that such a study of the first book of the *Elements* will prepare the reader for the more abstract and abstruse approaches that begin in Chapter 5.

4.1 Tacit, or Unstated, Assumptions

In Chapter 1 we saw five statements that Euclid chose as axioms. In addition to these, he assumed some statements that he regarded as common notions, statements such as: things equal to the same thing are equal; equals subtracted from equals are equal; and so on. These statements, the axioms and the common notions, along with quite a few definitions, are listed at the beginning of the *Elements*.[1] From all of these

[1] See Appendix, pp. 217ff.

listed assumptions Euclid proceeded to prove over four hundred theorems of which only the first twenty-eight shall be of direct concern to us.[1]

In studying the first twenty-eight theorems of the *Elements*, one finds that they more or less fall into two categories. There are those that may be regarded as theorems proper even in a system based on Hilbert's axioms. On the other hand, there are those that are merely proofs of the possibilities of certain constructions and proofs of existence via constructions. This second category, for the most part, will be of little use in the Hilbert system and we shall have no need to prove them as theorems.

The first three theorems of *Book 1* are of this latter type. Euclid justifies the construction of an equilateral triangle; he proves that one may mark off with a straightedge and collapsible compass a segment equal in length to a given segment; and, from a given point, draw a segment equal in length to a given segment. It is worthwhile to study these first three theorems. Not only do they have a certain literary charm, but they develop so nicely one after the other that they serve well as an introduction to the flavor of the work.

Present in the proofs of these three theorems are two *unstated* assumptions. This flaw of the *Elements* pervades the work. At times these assumptions are so subtly woven into the framework of the proofs that they pass unnoticed unless one knows exactly where to look; at other times they seem to have been introduced brazenly. In this instance they are of the first kind, and it is difficult indeed to discover on first reading what these unstated, or tacit, assumptions are. Consider Euclid's second theorem: (see Figure 4.1).

Theorem 2. To place at a given point (as an extremity) a straight line equal to a given straight line.

Proof: Let A be the given point, and BC the given straight line. Thus it is required to place at the given point A (as an extremity) a straight line equal to the given straight line BC.

From the point A to the point B let the straight line AB be joined (Post. 1); and on it let the equilateral triangle DAB be constructed. (Theorem 1).

Let the straight lines AE, BF be produced in a straight line DA, DB; (Post. 2); with center B and distance BC let the circle CGH be described; and again, with center D and distance DG let the circle GKL be described (Post. 3).

Then, since the point B is the center of the circle CGH, BC is equal to BG. Again, since D is the center of the circle GKL, DL is equal to DG. And in these DA is equal to DB; therefore, the remainder AL is equal to the remainder BG (C. N. 3).

But BC was also proved equal to BG; therefore, each of the straight lines AL, BC is equal to BG. And things which are equal to the same thing are also equal to one another; therefore, AL is also equal to BC (C. N. 1).

[1] See Appendix, p. 219.

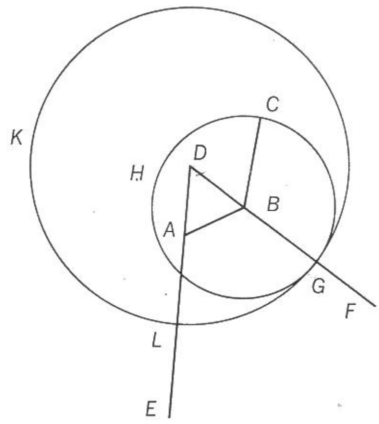

FIGURE 4.1

Therefore at the given point A the straight line AL is placed equal to the given straight line BC.

(Being) what it was required to do. ●

Can you find the flaw? Even a careful check of the proof of this theorem will show that all of the statements made in the proof actually follow from Euclid's axioms, common notions, definitions, or first theorem. The problem lies not in the statements *made* but in those that are *assumed but unstated.*

There are two such assumptions: that (under certain conditions) two circles intersect in two points and that, if a line passes through a point interior to a circle, it must intersect the circle (in two points).

In the circumstances given in the first three theorems there is not much doubt that these conditions hold true. But that is not the point. The criticism here is not that the assumptions are false; it is that the assumptions are not *stated.* If they are axioms, they should be stated as such; if they are theorems, they should be proven as such. But Euclid brings them in as hidden assumptions made credible by the diagrams. Nowhere does he make either an explicit assumption equivalent to these assumptions or one from which the truth of these statements would follow.

The rules must be such that this is not allowed in an axiomatic system. If a statement is not an axiom, definition, or theorem, it does not belong in a system. Here the solution is not difficult. One can introduce the statements themselves as axioms if one wishes or, better yet, can introduce a continuity axiom from which these statements would follow.

Hilbert chooses to introduce the Archimedean axiom to assure continuity of the lines and circles in his system. Stated rather intuitively, this axiom is:

Given two line segments, there exists an integral multiple of one which is greater than the other.

Another approach is to introduce the axiom of J. W. R. Dedekind (1831–1899), which states, more or less in Dedekind's own words:

If all points of a straight line fall into two classes, such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions.

We are going to refrain from introducing either of these axioms into our system because we feel that it is more important for the reader to see how many familiar theorems of ordinary plane geometry follow without an assumption of continuity. The main point is that if one wants continuity one must bring it in explicitly in some manner, not by unstated assumptions about diagrams. So agreed, let us turn to another problem.

In the proof of his famous fourth theorem, the *SAS* theorem, Euclid introduces for the first time the controversial method of *superposition.* It has been used and misused ever since and with much greater frequency than Euclid would have favored. Euclid himself does not use this technique very often but, evidently, saw no way to prove the fourth theorem without it.

There are various troubles involved in its use, all of which more or less boil down to the fact that in order to justify the use of superposition one must introduce such a statement as *figures remain unchanged when moved from one place to another.*

In fact, in the proof of the *SAS* theorem, Euclid brings in still another assumption; namely, that things which are equal may be made to coincide.

After years of controversy it was decided that it is impossible to use superposition without bringing in some axiom *justifying* superposition. Therefore, inasmuch as the use of superposition can be restricted to the proof of the *SAS* theorem, why not merely make that theorem an axiom? Such a course has been, and still is, often adopted.

To avoid this problem we shall, following Hilbert, introduce as an axiom a statement from which the *SAS* theorem readily follows without any such method as superposition. This involves introducing an undefined relation called *congruence.* This not only by-passes the need for superposition, it helps rid geometry of that much overused word "equals." Specifically, while Euclid will say that two triangles are *equal* if they coincide, he will also say that two triangles are *equal* if they have the same area (but perhaps cannot be made to coincide). There will be no discussion of area in this book; but, following Hilbert, we shall introduce the word "congruent" to replace the word "equal" as used in the first sense.

4.2 More Unstated Assumptions, Flaws, and Omissions

To show how many unstated assumptions Euclid introduces, and to illustrate other difficulties that arise in his system, let us examine and analyze the following elementary theorems from the *Elements:*

Theorem 5. In isosceles triangles the angles at the base are equal to one another; and, if the equal straight lines be produced further, the angles under the base will be equal to one another.
Proof: Let ABC be an isosceles triangle having the side AB equal to the side AC; and let the straight lines BD, CE be produced further in a straight line with AB, AC (Post. 2). I say that the angle ABC is equal to the angle ACB, and the angle CBD to the angle BCE.

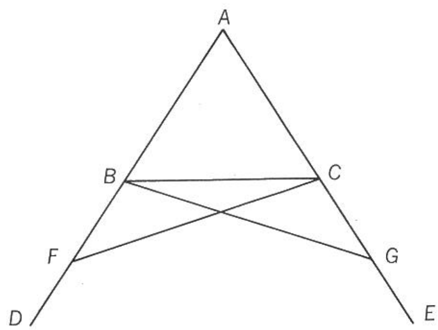

FIGURE 4.2

Let a point F be taken at random on BD; from AE the greater let AG be cut off equal to AF the less; and let the straight lines FC, GB be joined (Post. 1). Then, since AF is equal to AG and AB to AC, the two sides FA, AC are equal to the two sides GA, AB respectively; and they contain the common angle, the angle FAG. Therefore, the base FC is equal to the base GB, the triangle AFC is equal to the triangle AGB, and the remaining angles will be equal to the remaining angles respectively, namely, those which the equal sides subtend, that is, the angle ACF is equal to the angle ABG, and the angle AFC, to AGB (I.4).

And, since the whole AF is equal to the whole AG, and in these AB is equal to AC, the remainder BF is equal to the remainder CG.

But FC was also proved equal to GB; therefore, the two sides BF, FC are equal to the two sides CG, GB respectively; and the angle BFC is equal to the angle CGB, while the base BC is common to them; therefore, the triangle BFC is also equal to the triangle CGB, and the remaining angles will be equal to the remaining angles respectively; namely, those which the equal sides subtend. Therefore, the angle FBC is equal to the angle GCB and the angle BCF, to the angle CBG.

Accordingly, since the whole angle ABG was proved equal to the angle ACF, and in these the angle CBG is equal to the angle BCF, the remaining angle ABC is equal to the remaining angle ACB; and they are at the base of the triangle ABC.

But the angle FBC was also proved equal to the angle GCB, and they are under the base. Therefore, et cetera . . . Q.E.D. ●

There are three preliminary questions one might ask about this proof:

1. Is it true that angle BAC = angle FAC = angle GAB = angle FAG?
2. Must angle ABG be larger than and wholly contain angle CBG? Must angle ACF be larger than and wholly contain angle BCF?
3. If Question 2 is satisfied, is it satisfied in such a manner that it follows that the "remaining angle ABC is equal to the remaining angle ACB"?

One should certainly want the answers to the above questions to be yes; indeed, there is something wrong if the answer is anything but yes. But this is irrelevant in an axiom system. The question once again is not whether these statements are true but, rather, *whether the statements follow from other statements explicitly assumed or proved.* Unfortunately, the answer is no. The statements are "taken" as true. From the diagrams it is apparent that they are true, but this does not constitute a proof. If not axioms, and if not proved, statements should not be part of the proof, diagram or not.

Another proof of this theorem is more likely to appear in high school texts. Given the triangle ABC with equal sides AB and AC, bisect angle A.

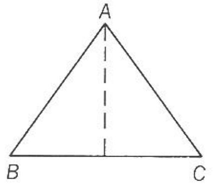

FIGURE 4.3

It follows easily from Theorem 4 that the two triangles are equal and hence that angles B and C are equal. As far as the *Elements*, and most high school texts, are concerned, this proof makes two unwarranted assumptions. It assumes that the bisector of an angle is unique and, furthermore, assumes that the bisector of an angle of a triangle intersects the opposite side. Euclid does not prove the former and ignores the latter, which he could not prove from his set of axioms even if he did consider it. Once again we emphasize that while these may be very obvious truths of geometry, the relevant question is whether or not they can be proved within the given system.

As another illustration of Euclid's rather easygoing introduction of unstated assumptions, consider the following theorem: (See Figure 4.4).

Theorem 6. If in a triangle two angles be equal to one another, the sides which subtend the equal angles will also be equal to one another.

Proof: Let ABC be a triangle having the angle ABC equal to the angle ACB; I say that the side AB is also equal to the side AC.

For, if AB is unequal to AC, one of them is greater.

Let AB be greater; and from AB the greater let DB be cut off equal to AC the less; let DC be joined.

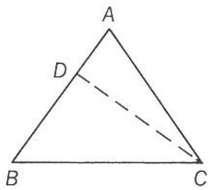

FIGURE 4.4

Then, since DB is equal to AC, and BC is common, the two sides DB, BC are equal to the two sides AC, CB respectively; and the angle DBC is equal to the angle ACB; therefore, the base DC is equal to the base AB, and the triangle DBC will be equal to the triangle ACB, the less to the greater; which is absurd.

Therefore, AB is not unequal to AC; it is therefore equal to it.

Therefore, et cetera . . . Q.E.D. ●

As in proofs throughout his work, Euclid is here assuming a trichotomy principle for segment lengths (later he will do the same for angle measure). In the third line of the proof he is assuming that given two segments AB and AC, either $AB = AC$; AB is less than AC; or AB is greater than AC, and exactly one of these conditions can hold.

In regard to the proof itself, it is apparently a proof by contradiction, but where exactly is the contradiction? He concludes that the "triangle DBC will be equal to the triangle ACB, the less to the greater, which is absurd." If he is not referring to *area*—and it is evident he is not—what can he mean when he says that "one triangle is less than another"? If there is a contradiction here (about triangles) it comes from the diagram, not from anything previously mentioned or proved in the system.

As a final illustration consider the following very important theorem (see Figure 4.5):

Theorem 16. In any triangle, if one of the sides be produced, the exterior angle is greater than either of the interior and opposite angles.

Proof: Let ABC be a triangle, and let one side of it BC be produced to D;

I say that the exterior angle ACD is greater than either of the interior and opposite angles CBA, BAC.

Let AC be bisected at E (I.10) and let BE be joined and produced in a straight line to F; let EF be made equal to BE (I.3), let FC be joined (Post. 1), and let AC be drawn through to G(Post. 2).

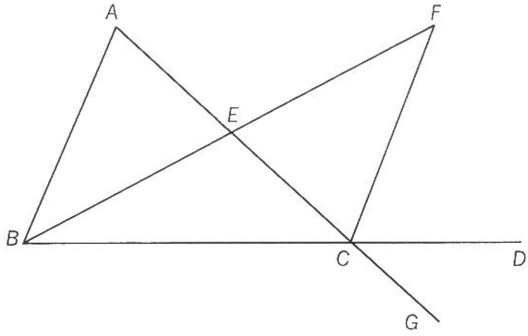

FIGURE 4.5

Then, since AE is equal to EC, and BE to EF, the two sides AE, EB are equal to the two sides CE, EF respectively; and the angle AEB is equal to the angle FEC, for they are vertical angles.

Therefore the base AB is equal to the base FC, and the triangle ABE is equal to the triangle CFE, and the remaining angles are equal to the remaining angles respectively, namely those which the equal sides subtend; (I.4).

Therefore the angle BAE is equal to the angle ECF. But the angle ECD is greater than the angle ECF; therefore the angle ACD is greater than the angle BAE.

Similarly also, if BC be bisected, the angle BCG, that is, the angle ACD (I.15), can be proved greater than the angle ABC as well.

Therefore, et cetera . . . Q.E.D. ●

A short digression is in order here. When Euclid asserts in his Axiom 2 that a straight line can be produced indefinitely, this does not in itself imply that straight lines are of infinite length. This point was ignored until 1854 when B. Riemann proposed that a distinction be made between a line being unbounded and a line being infinite in length.

As Euclid uses these terms, unboundedness is an *extent* concept; this means, as in the second axiom, that a line may be extended indefinitely in either direction. But unboundedness is compatible with the hypothesis that a line is infinite in length as well as with the hypothesis that a line is of finite length.

After this distinction was made there followed several new geometries in which straight lines are unbounded but not of infinite length; one such interpretation considers straight lines to be the great circles on the surface of a sphere. When "straight line" is so interpreted one gets geometries in which, while two points may determine a line, it may be that there are also an infinite number of lines through two points. Any two lines may enclose a space. This geometry is specifically "non-Euclidean" in the sense that there are no parallel lines whatsoever. Any of these considerations make for a "*non*-Euclidean" geometry.

Therefore, when Euclid tacitly assumes, in the proof of Theorem 16, that lines are of infinite length, the assumption is an important one to his geometry. Given the same proof for a triangle on a sphere it becomes invalid if BF is a semicircle, for then F falls on BD and angle ACD is equal to angle ECF and thus equal to angle BAC. And if BF is greater than a semicircle, angle ACD will be less than angle BAC. Thus, for a triangle on a sphere, not only does the proof fail but the "theorem" is a false statement; the theorems following from this one are false and there exists quite a different geometry. The geometry that does follow from the new assumptions is too complicated, unfortunately, to present here.

More relevant to this discussion is how Euclid knows, other than from looking at the diagram, that point F will fall in the interior of angle ACD. If a line is like a circle, it would be possible for F to fall between B and E. And even given that lines are of infinite length, how does one *prove* that F is in the interior of the angle so that it follows that angle ACF is less than angle ACD?

Euclid never defines or explicitly discusses such concepts as *interior* or *exterior*. He speaks again and again about one *side* or the other of a line but never defines what he means by a side of a line. And he ignores the concept of *betweenness*, so that whenever it is important to know that a point is between two others, or a ray is between two others (as when one wishes to say that one angle is less than another), he lets the diagram carry the weight of the argument.

In all these cases where Euclid tacitly introduces concepts by diagrams and uses diagrams to substantiate steps in a proof, he risks the chance of introducing not only false statements but self-contradictions. We shall try to show this now.

4.3 The Danger in Diagrams

Take a square each of whose sides is thirteen units and divide it into two rectangles of dimensions 8 by 13 and 5 by 13; then subdivide these into two congruent rectangular trapezoids whose parallel sides are 8 and 5 units long and into two congruent right triangles whose legs are 5 and 13 units long. Now rearrange these parts to form a rectangle. Finally, study Figure 4.6.

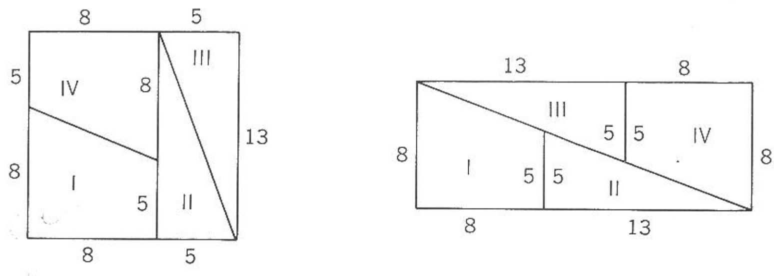

FIGURE 4.6

It can be seen more readily in the diagrams what is to be done than from the rather complicated set of directions preceding them. And this seems very simple until the following observation is made:

The area of the square is $13^2 = 169$ square units.

The area of the rectangle is $8 \times 21 = 168$ square units.

Obviously, something is wrong. It might help solve the dilemma to try this with different dimensions or with scissors and paper.

A more instructive illustration is the following well-known "paradox." We assume here that the reader has knowledge of the common elementary theorems of high school geometry.

All Triangles Are Isosceles

Given: *Any triangle ABC*

Construction: *Construct the bisector of angle A and the perpendicular bisector of BC, the side opposite angle A.*

Proof: Consider the following cases.

Case 1. The bisector of angle A and the perpendicular bisector of segment BC are either parallel or identical. In either case, the bisector of angle A is perpendicular to BC and hence, by definition, is an altitude. Therefore, the triangle is isosceles. (The conclusion follows from the Euclidean theorem: If an angle bisector and altitude from the same vertex of a triangle coincide, the triangle is isosceles.)

Suppose now that the bisector of an angle A and the perpendicular bisector of the side opposite are not parallel and do not coincide. Then they intersect in exactly one point, D. And there are three cases to consider:

Case 2. The point D is inside the triangle.
Case 3. The point D is on the triangle.
Case 4. The point D is outside the triangle.

For each case construct DE perpendicular to AB and DF perpendicular to AC, and in Cases 2 and 4 join D to B and D to C. In each case, the following proof now holds (see Figure 4.7):

$DE \cong DF$ because all points on an angle bisector are equidistant from the sides of the angle; $DA \cong DA$, and angle DEA and angle DFA are

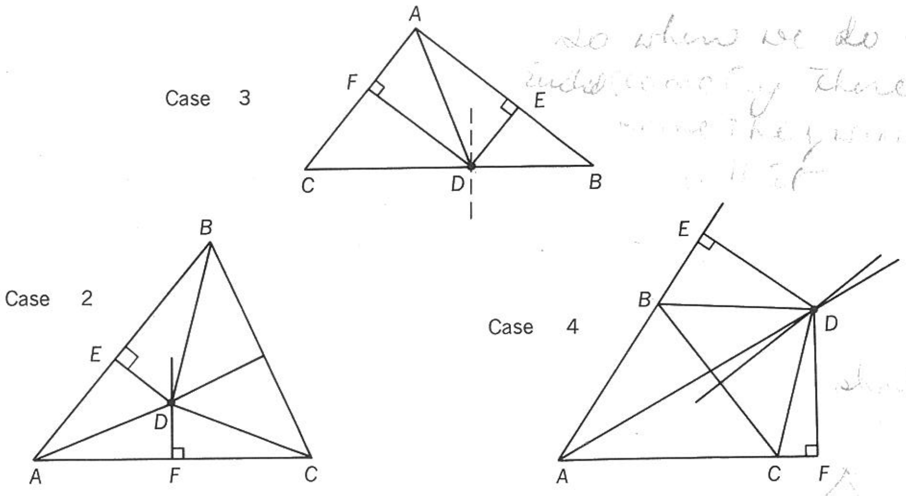

FIGURE 4.7

right angles; hence triangle ADE is congruent to triangle ADF by hypotenuse-leg theorem of Euclidean geometry. (We could also have used SAA theorem with $DA \cong DA$, and the bisected angle and right angles.) Therefore, we have $AF \cong AE$.

Now, $DB \cong DC$ because all points on the perpendicular bisector of a segment are equidistant from the ends of the segment. Also, $DE \cong DF$, and angles DEB and DFC are right angles. Hence triangle DEB is congruent to triangle DFC by hypotenuse-leg theorem. And hence $FC \cong BE$.

Therefore it follows that $AB \cong AC$—in Cases 2 and 3 by addition, in Case 4 by subtraction. Hence the triangle is isosceles.

To summarize the proof: It has been shown that if an angle bisector and the perpendicular bisector of the opposite side coincide or are parallel, the triangle is isosceles. It is known that if they neither coincide nor are parallel they must intersect in exactly one point which is either inside, on, or outside the triangle. But it has been shown that in every such case the triangle is isosceles. Therefore, every possible case has been considered and the theorem is proved.

Again, something is wrong. After a careful scrutiny, it will be obvious that not all cases have been considered. For while it is true that all locations of the intersection of the angle bisector and the perpendicular bisector of the opposite side have been considered, this has not been done for all locations of the points E and F. For example, in Cases 2 and 3, E and F might fall on the extension of AB and AC respectively, whereas in Case 4 E and F might fall *on* AB and AC respectively. It is left as an exercise to show that in these cases the proof still holds. Another possibility can easily be ruled out; namely, that E or F fall on a vertex of the triangle. But there are still other cases: the possibility that one of the points E or F might fall within a segment, while the other lies on an

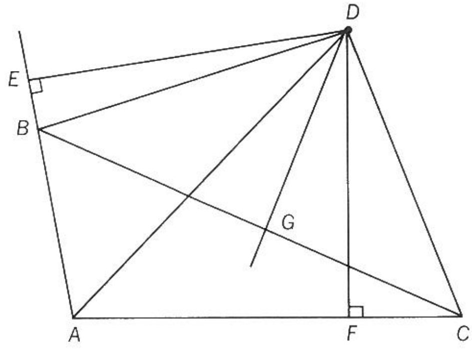

FIGURE 4.8

extension of its segment, has never been considered. Suppose this is done now: (See Figure 4.8).

Case 5. The same steps followed in Cases 1, 2, 3, and 4 may be repeated, but the results are not the same. For now it follows that $AB = AE - BE$ and $AC = AF + FC = AE + BE$ and since $AE - BE \neq AE + BE$, it follows that $AB \neq AC$. ●

It seems then that we have found the flaw in the reasoning: not every case had been considered. But, although this is true, has the flaw really been pointed out?

Most discussions of this paradox end at this stage. It has been shown that there are dangers in building an argument upon such weak foundations as diagrams; if that is all we were interested in showing, indeed we would be through with the problem. But there are broader aspects to the argument.

To appreciate fully the difficulties hidden here, let us approach the problem from a more logical perspective. In the proof of Case 1 a Euclidean theorem was cited. Its converse also holds; namely,

If a triangle is isosceles, then the angle bisector and the altitude from the same vertex coincide.

Thus the contrapositive must also hold; namely,

If the angle bisector and the altitude from the same vertex of a triangle do not coincide, the triangle is not isosceles.

The reader should not find it too difficult to prove that an angle bisector and altitude do not coincide if and only if the angle bisector and perpendicular bisector of the opposite side intersect in exactly one point. From this, it follows that *if the angle bisector and perpendicular bisector of the opposite side intersect in exactly one point, the triangle is not isosceles.*

But, it has just been proven by Cases 2, 3, and 4 that under the same hypotheses the triangle is isosceles.

This is a much more serious problem, one that is not answered by the mere existence of Case 5. In other words, it no longer suffices to show merely that Case 5 is a possible one; it must be shown, if no contradiction is to exist, that Cases 2, 3, and 4 are impossible.

This can be done as follows. It is a fact that the point D will always lie outside the triangle ABC and, of the two points F and E, one will always lie inside, the other outside the corresponding sides AC and AB of the triangle. For it can be proved that no line intersects all three sides of a triangle (unless one is extended) and it can be shown that the three points G (see Case 5), E, and F are collinear, so that one of the points E or F must lie on a side produced. If this were done, it would complete the argument and solve the "paradox."

What we wish to emphasize now is that any attempt to complete the above argument by the axioms and tools of Euclid's *Elements* will be hopeless. Not only does Euclid fail to present us with the tools, he does not even discuss such concepts as "between," "inside," or "outside." Within his system it is impossible to prove that a straight line cannot cut all three segments of a triangle. He must resort to argument from diagrams—and this, we should be more and more convinced, is a hazardous undertaking.

We can summarize some of the defects of the *Elements*, in particular those *tacit* or *unstated* assumptions which Euclid introduces into his work by means of diagrams or otherwise, as follows:

1. *continuity* of his figures
2. the *existence* of points and lines

3. *uniqueness* of certain points and lines
4. that a straight line which contains a vertex B and an interior point of a triangle ABC must also contain a point of the line segment AC
5. the existence of *order* relations on a line
6. the concepts of *inside* and *outside*

4.4 New Systems

To correct the defects listed in preceding sections, many axiom systems have been suggested and developed. Among these systems are those of Pasch——1882, Peano——1889, Hilbert——1899, Veblen——1904, and Birkhoff and Beatley——1940. Each is different; some have certain advantages over the others. Hilbert's system, perhaps because he is known as one of the outstanding mathematicians of the twentieth century, has had the most profound effect. Perhaps, too, this is because his system, as compared to the others, is most similar to Euclid's. Whatever the reasons, Hilbert's system has been so used, revised, and refined over the years that many variations of his system—changes in statements and phraseology—are now in existence. We shall use one such form as a basis, but shall incorporate many changes of our own.

Hopefully, the reader is prepared for this new approach to geometry. When we pause to prove something as obvious as the fact that a line has two "sides" or that a line "entering" a triangle must come "out," he will now be aware that it is not because we question the truth of these statements but that we wish to ascertain that they do indeed follow from the stated axioms.

Before we start, it should be made clear that we have specific objectives which require adopting the following restrictive procedures:

1. We shall introduce sufficient material to derive theorems related only to the first twenty-eight theorems of the *Elements*.
2. We shall refrain from introducing continuity considerations.
3. We shall not discuss any axioms needed for development of three dimensions, but shall, in effect, pretend that there are only two dimensions.

references

Dubnov, YA. S. *Mistakes in Geometric Proofs*. Translated by A. K. Henn and O. A. Titelbaum. Boston: D. C. Heath and Co., 1963.

Eves, Howard. *A Survey of Geometry*, vol. 1, chap. 8. Boston: Allyn and Bacon, Inc., 1963.

Fetisov, A. I. *Proof in Geometry*. Translated by T. M. Switz and L. Lange. Boston: D. C. Heath and Co., 1963.

Heath, T. L. *The Thirteen Books of Euclid's Elements*, 2d ed. New York: Cambridge University Press, 1926. Reprinted by Dover Publications, Inc., vol. 1., 1956.

Wilder, R. L. *Introduction to the Foundations of Mathematics*, chaps. 1 and 2. New York: John Wiley & Sons, Inc., 1952.

Young, J. W. *Lectures on Fundamental Concepts of Algebra and Geometry*, lectures IV & V. New York: The MacMillan Co., 1911.