



**HAL**  
open science

## Proof-checking Euclid

Michael Beeson, Julien Narboux, Freek Wiedijk

► **To cite this version:**

| Michael Beeson, Julien Narboux, Freek Wiedijk. Proof-checking Euclid. 2017. hal-01612807v1

**HAL Id: hal-01612807**

**<https://hal.science/hal-01612807v1>**

Preprint submitted on 8 Oct 2017 (v1), last revised 4 Nov 2018 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Proof-checking Euclid

Michael Beeson  
 Julien Narboux  
 Freek Wiedijk

October 3, 2017

## Abstract

We used computer proof-checking methods to verify the correctness of our proofs of the propositions in Euclid Book I. We used axioms as close as possible to those of Euclid, in a language closely related to that used in Tarski’s formal geometry. We used proofs as close as possible to those given by Euclid, but filling Euclid’s gaps and correcting errors. Euclid Book I has 48 propositions; we proved 213 theorems. The extras were partly “Book Zero”, preliminaries of a very fundamental nature, partly propositions that Euclid omitted but were used implicitly, partly advanced theorems that we found necessary to fill Euclid’s gaps, and partly just variants of Euclid’s propositions. We wrote these proofs in a simple fragment of first-order logic corresponding to Euclid’s logic, debugged them using a custom software tool, and then checked them in the well-known and trusted proof checkers HOL Light and Coq.

## 1 Introduction

Euclid was the “gold standard” of rigor for millenia. The *Elements* of Euclid set the standard of proof used by Isaac Newton in his *Principia* and even Abraham Lincoln claimed to have read all ten books of Euclid and learned from it how to prove something in court. The *Elements* also inspired the form of the American Declaration of Independence. In the modern era, beginning already in the nineteenth century, the standards of proof in mathematics became more demanding, and the imprecisions and gaps in Euclid were more apparent than before. Even before that time, some mathematicians focused on the perceived flaw that the Fifth Postulate (the “parallel postulate”, or “Euclid 5”) was less intuitively evident than the other four.<sup>1</sup> Efforts to remove this “flaw” by proving Euclid 5 led to the development of non-Euclidean geometry, showing that in fact Euclid 5 was a necessary postulate, not a flaw.

---

<sup>1</sup>Proclus [33], writing in the fifth century, said that Euclid 5 needed a proof, and tried to supply one; and he was not the first, since he also criticized a previous attempt by Ptolemy. See [21], Ch. 5 for further history.

Nevertheless there are flaws in Euclid, and we want to discuss their nature by way of introduction to the subject.<sup>2</sup> The first gap occurs in the first proposition, I.1, in which Euclid proves the existence of an equilateral triangle with a given side, by constructing the third vertex as the intersection point of two circles. But why do those two circles intersect? Euclid cites neither an axiom nor a postulate nor a common notion. This gap is filled by adding the “circle–circle” axiom, according to which if circle  $C$  has a point inside circle  $K$ , and also a point outside circle  $K$ , then there is a point lying on both  $C$  and  $K$ .

There is, however, a second gap in the proof of I.1. Namely, the proposition claims the existence of a *triangle*, which by definition is (or at least is determined by) three *non-collinear* points.<sup>3</sup> Why are the three points in question not all on a line? Intuitively, if they were, then one would be between the other two, violating Euclid’s common notion that the part is not equal to the whole, since all the sides are equal.

A formal proof of Prop. I.1 cannot follow Euclid in ignoring this issue. The proposition that if  $B$  is between  $A$  and  $C$  then  $AB$  is not equal to  $AC$  therefore must precede proposition I.1, unless one is to consider it as an axiom formalizing one of Euclid’s common notions. In the next section, we discuss the axioms and postulates of Euclid, and how we have chosen to render them in modern first-order form.

The two gaps in I.1 illustrate two common failures. Many of the gaps in Euclid fall into one of these categories:

- (i) A failure to prove that a point clearly shown in the diagram actually exists, e.g. that two lines really do intersect, or as in I.1 two circles.
- (ii) A failure to prove that points shown in the diagram to be non-collinear, are in fact non-collinear.
- (iii) A failure to prove that a point shown in the diagram to be between two other points, is in fact between those points.

Another example of an error of type (i) is in the second paragraph of Prop. I.44, “let  $FG$  be drawn through to  $H$ ”. Here  $H$  has not been proved to exist, a strange omission in that a few lines later Euclid does feel the need to use Postulate 5 to prove that  $K$  exists; but then two lines later “let  $HA, GB$  be produced to the points  $L, M$ ”. That is, the lines shown as intersecting at  $L$  and  $M$  do in fact intersect—but Euclid offers no justification for that line of the proof. There are dozens of such fillable gaps in Euclid’s proofs, some more easily filled than others.

Not every error in Euclid falls into these categories, however. Consider Prop. I.6, in which Euclid bisects an angle. The method is to lay off the same length on both sides of the angle, and then construct an equilateral triangle on the base thus formed. Connecting the vertex of the original angle with the vertex of the equilateral triangle, we get the angle bisector.

---

<sup>2</sup>It is customary to refer to the propositions of Euclid with notation like I.44, which means Proposition 44 of Book I.”

<sup>3</sup>Euclid never defines *triangle*, although he does define *right triangle* and *equilateral triangle*. Instead he mentions *trilateral figure*, which is “contained by three straight lines.” We read that to imply non-collinearity, for otherwise there would be only one line, and nothing “contained.”

Oops, but the two points might coincide! Well, in that case we ought to have drawn the *other* equilateral triangle, on the other side of the base. But Prop. I.1 did not provide for the construction of two triangles, and we cannot easily construct “the other one.” We certainly need to expand the list of ways in which Euclid’s proofs fall short by at least two more items:

(iv) A failure to prove that points shown in the diagram to be distinct are in fact distinct.

(v) A failure to show that points are on the same side (or opposite sides) of a line, as the diagram shows them to be.

Even if we could solve these problems, the proof of I.6 still would not be correct, since we would still need to show that the angle bisector constructed does in fact lie in the interior of the angle. That difficulty brings us to an important point. There is no “dimension axiom” in Euclid to guarantee that we are doing plane geometry. Hence “circles” are really “spheres”, and rather than just two equilateral triangles on a given base, there are infinitely many. So even if the vertex of the equilateral triangle is distinct from the vertex of the original angle, why does the line between them lie in the interior of the angle? In fact that is a problem even in two dimensions—an example of (iii) above.

Even if it were possible to fix that problem by adding a dimension axiom, that would not be desirable. Euclid didn’t just forget to write down a dimension axiom. In Book X and beyond, Euclid works in three-space, and the culmination of the whole series of books is the study of the Platonic solids. Hence it is clear that Euclid is not meant to be restricted to plane geometry. In the absence of a dimension axiom, it is good advice to the reader to visualize “circles” as spheres. Then the two circles used in I.1 have not just two but many intersection points. The circle–circle axiom only guarantees the existence of *one* intersection point.<sup>4</sup> Therefore we conclude that Euclid’s proof of I.6 is fatally flawed. We prove it another way: by first showing how to bisect a line, and then using that to bisect an angle. The simple and ingenious proof that a line can be bisected if it is the base of an isosceles triangle was apparently not noticed until Gupta’s thesis in 1965 [23], but could very well have been discovered by Euclid.

The fact that there is no dimension axiom is not always remembered in Book I, and is the source of several difficulties. A good example is Prop. I.7, which says that if  $ABC$  is a triangle and  $ABD$  is another triangle congruent to  $ABC$  (that is,  $AC = AD$  and  $BC = BD$ ), and  $C$  and  $D$  are on the same side of  $AB$ , then  $C$  is equal to  $D$ . Look at the figure for I.7; in two dimensions it appears contradictory, but as soon as you think that it might be in three dimensions, the contradiction disappears. The statement of Prop. 7 does correctly include the hypothesis that  $C$  and  $D$  are on the same side of  $AB$ , but Euclid never uses that hypothesis in the proof. That is not surprising, since he never defined “same side”, so he

---

<sup>4</sup>Since Euclid did not state the axiom, one could consider strengthening it to state the existence of an intersection point on a given side of the line connecting the centers of the circles. But Euclid also did not say in I.1 or I.6 anything about this problem; and we *are* able to prove I.6 from the more fundamental form of circle–circle.

had no possible way to use that hypothesis. It should have been used to verify the claim that angle  $DCB$  is less than angle  $DCA$ , because there is a point of intersection of  $AD$  and  $CB$ . This turns out to not be provable, even after proving a number of more basic propositions in “Book Zero”; we could not prove I.7 without using “dropped perpendiculars”, which are only constructed much later in Euclid.

A well-known geometer told me “there are no errors in Euclid”, in the sense that the statements of all the propositions are true in the plane. If we supply Tarski’s definition of “same side”, an even stronger version of that claim is true: the statements of all the propositions are true in every finite-dimensional space  $\mathbb{R}^n$ . However, the same cannot be said of the *proofs*. Many of these have problems like those of I.6 and I.7; that is, we could fix these problems only by proving some other propositions first, and the propositions of the first half of Book I had to be proved in a different order, namely 1,3,15,5,4,10,12,7,6,8,9,11, and in some cases the proofs are much more difficult than Euclid thought. After proving those early propositions, we could follow Euclid’s order better, and things went well until Prop. 44. In Propositions 44 and 47 there are numerous points of difficulty, which took extra propositions to resolve. As one example we mention the proposition that every square is a parallelogram, which Euclid uses implicitly in proving Prop. 47. (By definition a square has equal sides and four right angles, and a parallelogram is a quadrilateral with opposite sides parallel.) Euclid could and should have proved that.

The aims of this paper are as follows:

(i) Fix Euclid’s axioms (and common notions), using an axiom system rather close to Euclid’s, but including axioms about betweenness that Euclid omitted, and with other changes discussed below.

(ii) Give correct proofs of all the propositions in Book I from the new axioms, following Euclid’s proofs as closely as possible.

(iii) Show that those proofs are indeed correct by checking the proofs using the proof-checking programs HOL Light and Coq.

(iv) Show that the axioms are indeed correct by computer-checking proofs that the axioms hold in the Cartesian plane  $\mathbb{R}^2$ .<sup>5</sup>

Accordingly, in this paper we limit the discussion of geometry to the description of the axioms, the description of a few specific flaws in Euclid’s reasoning, and a discussion of Euclid’s notion of “equal figures” that is necessary to verify the axioms we use about that notion. Our focus in this paper is on proof checking. What we report on here would still have been worth doing, even if there were no gaps or errors in Euclid. The details of our proofs, and a discussion of the errors in the original proofs of Euclid, will be published separately, with a focus on the geometry.

The formal proofs themselves, as well as the PHP and ML scripts with which we manipulated and debugged these proofs, will be available on the Web by Oct. 10, 2017. Look for links to them at

<http://www.michaelbeeson.com/research/CheckEuclid/index.php>

---

<sup>5</sup>The axioms hold in  $\mathbb{F}^2$ , where  $\mathbb{F}$  is any Euclidean field (an ordered field in which positive elements have square roots), but we have not computer-checked a proof of that fact.

A scholarly history of the previous attempts to axiomatize Euclid would require a long paper in itself; we offer only a few highlights here. The best known attempt is Hilbert's 1899 book [25]. Hilbert had been a vocal proponent of the axiomatic method, and his book was probably meant partly to illustrate that method on the example of geometry. First-order logic was in its infancy and Hilbert's system was not first order. He made use of Archimedes's axiom, and his continuity "axiom" is a strange mixture of logic and model theory. The fundamental idea to use betweenness and congruence as primitive relations goes back to Pasch[30]. Further contributions by Mollerup [28], Veronese [39], Pasch [30], and Peano[31] are discussed below in connection with the axioms they helped to develop. After Hilbert, the most important work is the axiom system of Tarski. This is a first-order system, and not only is it first-order, it is points-only, meaning that there are variables only for points. Lines are given by two points, and angles by three points, and equality between angles is a defined notion. This fits Euclid very well: Euclid almost always refers to lines by two points, and angles by three points. Tarski's system was developed in 1927, but publication was long delayed; for the history see the introduction to (the Ishi press edition of) [35]. Although a development of Euclid in Tarski's system could have been done in the 1960s at Berkeley, it was not done. Instead efforts focused on reducing the number of Tarski's axioms by finding dependencies, and on proving fundamental results like the existence of perpendiculars and midpoints without using the parallel postulate or any continuity axioms. The results of these efforts finally appeared in [35], which contains the remarkable results of [23].

It is now 34 years since the publication of [35], and meantime, the technology of proof checking by computer has advanced. Our predecessors stopped at the threshold, so to speak, by working on neutral geometry and minimal axioms systems, instead of formalizing Euclid. We serendipitously find ourselves in the situation where it is possible for us to take up that task, and also to verify (using existing computer proof checkers) that our proofs are flawlessly correct.

## 2 Language

Euclid did understand the fundamental point that not every fact can be proved; the first fact accepted could not be proved because there would be nothing to prove it from. But he did not understand that similarly, not every concept can be defined. Thus he famously attempted to define "point" and "line" and "rectilinear angle." (The Greeks also considered "angles" formed by curved sides.) These natural-language "definitions" are not mathematical usable; so in practice Euclid treated points, angles, circles, and lines all as primitive notions. In addition, Euclid treated "figures" as a primitive notion. A figure seems to be what we would now call a simple closed polygon; in Book I, only triangle and quadrilaterals are used. Euclid also accepted concepts of "equal" and "greater than" for each of points, angles, circles, lines, and figures without definition. In the middle of the nineteenth century, it was recognized that "betweenness" and

“equidistance” were good primitives for geometry, and later it was realized that it is possible to work with angles represented by triples of points, instead of taking them as primitive, so angle equality and inequality are defined concepts. That is what we do in our formal work. Thus all our axioms, except circle–circle, are formulated in a “points-only” language, in which the fundamental relations are betweenness and equidistance.

Betweenness is a 3-ary relation  $\mathbf{B}(A, B, C)$ , which Euclid wrote as “ $B$  lies on the finite straight line  $AC$ .” We interpret that as *strict* betweenness, i.e. the endpoints do not lie on the line.<sup>6</sup> On the occasions when a betweenness statement had to be proved (e.g. in Prop. I.14), instead of saying “ $B$  is between  $C$  and  $D$ ”, Euclid said “ $CB$  is in a straight line with  $BD$ .” So he did understand the concept.

Collinearity is the relation  $L(A, B, C)$  expressing that either two of the points are equal or one lies between the other two. This is a statement about points only. It seems that for Euclid, lines were primitive objects, rather than sets, and the incidence relation (point lies on line) too fundamental even to notice, as it does not occur in the list where “point” and “line” are “defined.” Whether close to Euclid or not, we use the first-order formulations of betweenness (a primitive) and collinearity (defined).<sup>7</sup>

“Equidistance” is a 4-ary relation best thought of as congruence of finite lines, “ $AB$  is equal to  $CD$ ”. Euclid, or at least his translator Heath, used “equal” rather than “congruent”. Inequality of finite lines  $AB < CD$  is defined as  $AB$  is equal to  $CE$  for some  $E$  between  $C$  and  $D$ . Then “the whole is not equal to the part” (for lines) becomes  $AB < CD$  implies  $AB$  is not equal to  $CD$ , which can be proved. Let us ask, which of Euclid’s common notions must we retain as axioms, and which can we prove? The answer: all we need to retain as common notions are the equality axioms (for equality of points), the symmetry, reflexivity, and transitivity of congruence, and the axiom that  $AB$  is equal to  $BA$ .

There is one exception to our “points-only” approach. In order to follow Euclid more closely, we allow giving a name to a circle. Circles are given by point and radius, so we can say “let  $J$  be the circle of center  $P$  and radius  $AB$ .” (Here the “radius” is not a number but a finite line.) Then “inside” and “outside” can be defined using inequality of finite lines, and the circle–circle continuity axiom can be translated straightforwardly from the informal English version given above. To fit our theory into the framework of first-order predicate calculus, we use a two-sorted predicate calculus, one sort for points and one for circles. with relations “on” and “inside” and “outside” that take both a point argument and a circle argument.

Euclid never gives circles a single-character name as we do. Nor does he name circles by center and radius, or center and point-on-circle. Instead he names circles by listing three points that lie on the circle. One of those

---

<sup>6</sup>Hilbert used strict betweenness; Tarski used non-strict betweenness, on purpose because the degenerate cases could be used to reduce the number of axioms. We use strict betweenness, on purpose to avoid degenerate cases that express unintended things and have to be separately worried about.

<sup>7</sup>Most of the time Euclid’s lines are finite, which may confuse a modern reader at first, since today finite lines are called “segments”, and “line” means “infinite line.” Hilbert made (infinite) lines primitive objects, but treated finite lines (segments) as sets of points.

points is a point that appears to exist in the diagram, and is conjured into existence without proof by the act of naming it. This naming technique papers over the lack of the circle–circle axiom in Euclid, and introduces a gap into the proof every time it is used. We therefore *must* deviate from Euclid’s naming convention for circles.

Betweenness and equidistance are sufficient as primitive relations for geometry, but the latter part of Euclid Book I uses another primitive relation, “equal figures”, which is discussed in §5. We mention it here only to note that the complete definition of our language requires inclusion of the primitive relations discussed in §5.

### 3 Definitions

Euclid gives a long list of definitions at the beginning of Book I. We do the same. Euclid’s list has some important omissions, notably “same side” and “opposite side”. These are defined in Fig. 1.<sup>8</sup> Euclid defined “square” but omitted “parallelogram” and “rectangle”. He defined “parallel lines” to be lines that do not meet but lie in the same plane (thus illustrating that his omission of a dimension axiom was no accident!) On the other hand, he failed to define “lies in the same plane”. Once we have defined “same side” as in Fig. 1, it is easy to define “lies in the same plane”, as each line and point not on the line determine two half-planes, together making up a plane. In the formal statement of “same side”, a line is specified by two distinct points  $p, q$ , and the incidence of  $x$  on that line is expressed by “ $p, q, x$  are collinear”. Formally we use the relation  $L(p, q, x)$  defined above in terms of betweenness. This definition exemplifies how one works with points only, avoiding the explicit mention of lines. The price one has to pay for this simplification is that one then has to prove that it doesn’t matter which particular points  $p, q$  we chose to represent the line. That is, “same side” is invariant if  $p$  and  $q$  are changed to some other pair of distinct points each of which is collinear with  $p$  and  $q$ .

Euclid’s failure to define “lies in the same plane” leaves us to complete his definition of “parallel”. First, we discuss the “not meeting” part of the definition. Lines in Euclid are finite, but “parallel” is about infinite lines. So “ $AB$  does not meet  $CD$ ” means that no matter how those two finite lines are produced, the lengthened lines still do not have a point in common. In other words, there is no point collinear with both  $AB$  and  $CD$ . On the other hand, “ $AB$  crosses  $CD$ ” means that there is a point both between  $A$  and  $B$ , and between  $C$  and  $D$ .

We define “Tarski-parallel” by “ $AB$  and  $CD$  do not meet, and  $C$  and  $D$  lie on the same side of  $AB$ .” This is clearly not what Euclid intended, as to Euclid it seems obvious that if  $AB$  is parallel to  $CD$  then  $CD$  is parallel to  $AB$ . So we define instead that  $AB$  is parallel to  $CD$  if there is no point collinear with both  $AB$  and  $CD$ , and there are four points

---

<sup>8</sup>This definition is due to Tarski [35]. Hilbert had planes as a primitive concept, and discussed “same side” and “opposite side” only in the context of a fixed plane, using a definition that would not work without having planes as a primitive concept. Tarski’s definition of “same side” is vital for making possible a points-only formalization that would work in more than two dimensions.



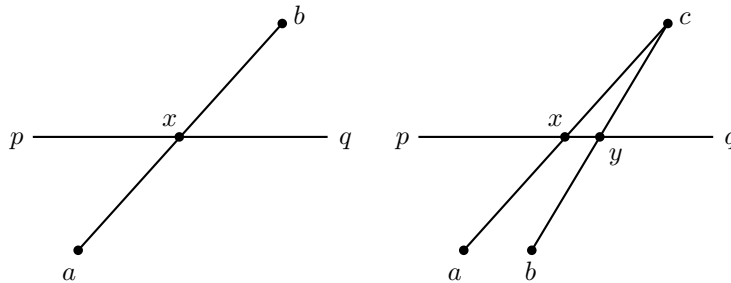


Figure 1: (Left)  $a$  and  $b$  are on the opposite side of  $pq$ . (Right)  $a$  and  $b$  are on the same side of  $pq$  if there exist points  $x$  and  $y$  collinear with  $pq$ , and a point  $c$ , such that  $\mathbf{B}(a, x, c)$  and  $\mathbf{B}(b, y, c)$ .

$a, b, c, d$  with  $a$  and  $b$  collinear with  $AB$ , and  $c, d$  collinear with  $CD$ , and  $ad$  crosses  $bc$ . With this definition, there is a very short proof of the symmetry property. On the other hand, the two definitions can be proved equivalent. It follows that if  $AB$  and  $CD$  are parallel then  $A$  and  $B$  are on the same side of  $CD$ , which is quite often actually necessary, but never remarked by Euclid.

Euclid defines a square to be an quadrilateral with four right angles in which all the sides are equal. He does not specify that all four vertices lie in the same plane. This is not trivial to prove, but we did prove it, so Euclid’s definition does not require modification. Euclid does not define “rectangle”. One would like to define it as a quadrilateral with four right angles. It is a theorem that such a figure must lie in a plane. However, the proofs we found involve reasoning “in three dimensions”. Even though Euclid Book I has no dimension axiom, and we must therefore be careful not to assume one, nevertheless all the *proofs* in Book I deal with planar configurations. We therefore define “rectangle” to be a quadrilateral with four right angles, whose diagonals cross, that is, meet in a point. This condition is one way of specifying that a rectangle lies in a plane. We can then prove that a rectangle is a parallelogram.

The Appendix contains a complete machine-generated list of our definitions.

## 4 Angles

We take only points and circles as primitive objects. Angles are treated as ordered triples of non-collinear points,  $ABC$ . The point  $B$  is the vertex of the angle. Equality of angles is a 6-ary relation, which we write informally as “angle  $ABC = abc$ ”. The definition is that there exist four points (one on each side of each angle) that form, with the vertices  $B$  and  $b$ , two congruent triangles. (Two triangles are congruent, by definition, if all three pairs of corresponding sides are equal.)<sup>9</sup> This definition does

<sup>9</sup>This is not the same definition as used in [35], but it works, and seems simpler to us; perhaps the one in [35] seems simpler in the presence of function symbols for line extension.

not permit “straight angles”, “zero angles”, or angles “greater than 180 degrees.” Such “angles” are also not used in Euclid.<sup>10</sup>

A point  $F$  lies in the interior of angle  $ABC$  if it is between two points lying on the two sides of the angle. Angle ordering is defined by  $abc < ABC$  if angle  $abc$  is equal to angle  $ABF$ , for some  $F$  in the interior of  $ABC$ . Note that these definitions make sense without any dimension axiom; that is, they work fine in three-space.

We then have to prove as theorems those properties of angle equality and ordering that Euclid assumed as “common notions”: reflexivity, symmetry, and transitivity of angle equality; the fact that angle  $ABC$  is equal to angle  $CBA$ ; transitivity of angle ordering. The fact that an angle cannot be both equal to and less than the same angle is quite difficult to prove, although taken for granted by Euclid in several proofs. That is, of course, the key result needed to prove antisymmetry and trichotomy for angle ordering.

Hilbert [25] took angles as primitive, and had an axiom about copying angles that specified the *uniqueness* of the copied angle. The uniqueness assumption builds in as an axiom the property that an angle cannot be both less than and equal to itself. Since this can be proved, it might be considered an imperfection to assume it as an axiom. While Hilbert took angles and equality of angles as primitive, he did define angle ordering just as we do. Because of the uniqueness part of his angle-copying axiom, he had no difficulty proving trichotomy.

## 5 Axioms and Postulates

Euclid had three groups of what would now be called axioms: common notions, axioms, and postulates. The common notions were intended to be principles of reasoning that applied more generally than just to geometry. For example, what we would now call equality axioms. The axioms and postulates were about geometry. The distinction between an “axiom” and a “postulate”, according to Proclus [33], p. 157, is that a postulate asserts that some point can be constructed, while an axiom does not. In modern terms an “axiom” is purely universal, while a postulate has an existential quantifier.

Heath’s translation lists five common notions, five postulates, and zero axioms. Simson’s translation [17] lists three postulates, twelve axioms, and zero common notions. The extra axioms are discussed by Heath on p. 223 of [18], where they are rejected.

Euclid’s first common notion is “things equal to the same thing are equal to each other.” That is,

$$a = c \wedge b = c \rightarrow a = b.$$

Modern mathematicians would prefer

$$a = c \wedge c = b \rightarrow a = b.$$

---

<sup>10</sup>Except perhaps in Propositions 13 and 17, which speak about pairs of angles being equal to other pairs of angles. We did not formalize those two propositions, as it would have required defining angles greater than 180°, and those propositions are never used in the rest of Book I.

But then, they need symmetry as a separate axiom ( $a = b \rightarrow b = a$ ), while that can be proved from Euclid's axiom above. We follow Euclid in this matter, although of course it is of no serious consequence. We do need reflexivity, the formalization of the common notion that anything is equal to itself:  $a = a$ .

Equality also enjoys the substitution property for each predicate in our language:

$$\mathbf{B}(a, b, c) \wedge a = A \wedge b = B \wedge c = C \rightarrow \mathbf{B}(A, B, C)$$

$$ab = cd \wedge a = A \wedge b = B \wedge c = B \wedge d = D \rightarrow AB = CD$$

and similarly for the predicates for "equal figures." In practice, the proofs are checked assuming the second-order property:

$$a = b \rightarrow \forall P (P(a) \iff P(b))$$

which allows to avoid introducing a separate axiom for each predicate. That is, we do not actually use the substitution axioms for individual predicates, but allow the substitution of  $A$  for  $B$  in any derived formula, when  $A = B$  or  $B = A$  has been derived.

Of Euclid's common notions as listed in the Heath translation, we keep the first, reject the fourth, and prove the other three as theorems. (The fourth is used to prove I.4, the SAS principle, by superposition, a proof we also reject.)

We include as common notions the axioms that say that congruence of finite lines is an equivalence relation, and the principle that  $AB$  is congruent to  $BA$ . A complete machine-generated list of the common notions is in the Appendix.

## Betweenness Axioms

Euclid never made explicit mention of betweenness, ignored all places where it should have been proved, and had no axioms for proving betweenness statements. We will not discuss the historical origins of the following axioms, nor the possibilities for reducing their number (this is certainly not a minimal set, but the proofs required to eliminate some of these axioms are long and difficult.) We give them the names they are given in our formal development, which is why there are no spaces in those names.

$\neg \mathbf{B}(a, b, a)$	<b>betweennessidentity</b>
$\mathbf{B}(a, b, c) \rightarrow \mathbf{B}(c, b, a)$	<b>betweennesssymmetry</b>
$\mathbf{B}(a, b, d) \wedge \mathbf{B}(b, c, d) \rightarrow \mathbf{B}(a, b, c)$	<b>innertransitivity</b>

The following axiom is called **connectivity** and can be rendered in English as "If  $B$  and  $C$  lie on the finite straight line  $AD$ , and neither is between  $A$  and the other, then they are equal."

Formally:

$$\mathbf{B}(a, b, d) \wedge \mathbf{B}(a, c, d) \wedge \neg \mathbf{B}(a, b, c) \wedge \neg \mathbf{B}(a, c, b) \rightarrow b = c$$

This principle was expressed in antiquity as “a straight line cannot enclose an area.”<sup>11</sup> It is closely related to the principle known in modern times as “outer connectivity”, which says that if line  $AB$  has two extensions  $C$  and  $D$  then either  $C$  is between  $B$  and  $D$  or  $D$  is between  $B$  and  $C$ . We prove outer connectivity as a theorem from the connectivity axiom.<sup>12</sup>

## Null lines

A “null line” or “empty line” is one with equal endpoints,  $AA$ . There are two axioms about null lines:

$ab = cc \rightarrow a = b$     **nullsegment1**, only a null line is equal to a null line.  
 $aa = bb$     **nullsegment2**, all null lines are equal.

Euclid never considered null lines, but we cannot dispense with either of these axioms.

## Extension of lines

Euclid postulated that every line can be extended, but (at least in Heath’s translation) did not say by how much.<sup>13</sup> Instead we postulate that every line  $AB$  can be extended by the amount  $CD$ ; that is, there exists a point  $E$  such that  $\mathbf{B}(A, B, E) \wedge BE = CD$ . Since Euclid never deals with null lines, we require both  $A \neq B$  and  $C \neq D$  as hypotheses.

Euclid’s Prop. I.2 asserts that given any point  $A$  and line  $CD$  there is a point  $E$  with  $AE = CD$ . The extension postulate renders Prop. I.2 superfluous. That is a matter of some regret, since I.2 has a beautiful proof. We did investigate weakening the extension postulate so that I.2 would require a proof, but then we need additional betweenness axioms. In addition, after I.2 is proved we then need the line–circle axiom to prove the extension postulate, which is “overkill.” We therefore decided to use the extension postulate as described above, rather than attempt to weaken it.<sup>14</sup>

---

<sup>11</sup> It is discussed by Proclus [33], p. 126, who thinks it superfluous as it is included in the meaning of Postulate 1. Apparently Simson was not convinced, as his translation [17] lists it as an axiom. Heath rejects it as an axiom (p. 232 of [18]), not on mathematical grounds, but because he came to the conclusion that it is an “interpolation”, i.e., is not in the original Euclid, in spite of being included in three of the “best manuscripts.”

<sup>12</sup>Outer connectivity was discussed already by Proclus, who stated it as “two straight lines cannot have a common segment” [33], p. 168-9, §216-17. Proclus says it is implicit in Euclid’s line extension axiom. Nevertheless, Proclus considers some possible proofs of it—but not the ingenious proof offered by Potts in the commentary to Prop. I.11 in [32], p. 14, which shows that outer connectivity follows from perpendiculars and the fact that an angle cannot be less than itself. The latter, however, is a difficult theorem, if it (or a close equivalent) is not assumed as an axiom.

<sup>13</sup>The Simson translation [17] renders the extension postulate as “*That a terminated straight line may be produced to any length in a straight line.*” Perhaps Euclid’s extension postulate said more than Heath’s translation indicates.

<sup>14</sup>This decision is inconsequential; if we weaken the extension postulate, add more betweenness axioms, prove Prop. I.2 and use the line–circle axiom to prove the omitted extension axiom, then from that point on the development is the same anyway. The only effect is to save the beautiful proof of I.2.

## Five-line Axiom

Euclid attempted, in Proposition I.4, to prove the side-angle-side criterion for angle congruence (SAS). But his “proof” appeals to the invariance of triangles under rigid motions, about which there is nothing in his axioms, so for centuries it has been recognized that in effect SAS is an axiom, not a theorem.

Instead of SAS itself, we take an axiom known as the “five-line axiom.” This axiom is illustrated in Fig. 2. Its conclusion is, in effect, the congruence of triangles  $dbc$  and  $DBC$  in that figure. Its hypothesis expresses the congruence (equality, in Euclid’s phrase) of angles  $dbc$  and  $DBC$  by means of the congruence of the exterior triangles  $abd$  and  $ABD$ .

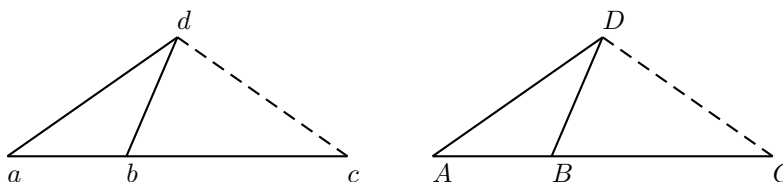


Figure 2: If the four solid lines on the left are equal to the corresponding solid lines on the right, then the dotted lines are also equal.

Our version of the five-line axiom was introduced by Tarski, although we have changed non-strict betweenness to strict betweenness.<sup>15</sup>

## Pasch’s Axiom

Pasch [30] introduced the axiom that bears his name, in the form that says that if a line enters a triangle through one side, it must exit through another side (or vertex). That version, of course, is only true in a plane. Seven years later, Peano [31] introduced what are now called “inner Pasch” and “outer Pasch”, which work without a dimension axiom.<sup>16</sup> See Fig. 3. In that figure, we use the convention that solid dots indicate points assumed to exist, while an open circle indicates a point that is asserted to exist.

Technically “Pasch’s axiom” should be “Pasch’s postulate”, since it makes an existential assertion, but the terminology is too well-established

<sup>15</sup>The history of this axiom is as follows. The key idea (replacing reasoning about angles by reasoning about congruence of segments) was introduced (in 1904) by J. Mollerup [28]. His system has an axiom closely related to the 5-line axiom, and easily proved equivalent. Tarski’s version [38], however, is slightly simpler in formulation. Mollerup (without comment) gives a reference to Veronese [39]. Veronese does have a theorem (on page 241) with the same diagram as the 5-line axiom, and closely related, but he does not suggest an axiom related to this diagram.

<sup>16</sup>Axiom XIII in [31] is outer Pasch, with  $\mathbf{B}(a, b, c)$  written as  $b \in ac$ . Axiom XIV is inner Pasch. Peano wrote everything in formal symbols only, and eventually bought his own printing press to print his books himself.

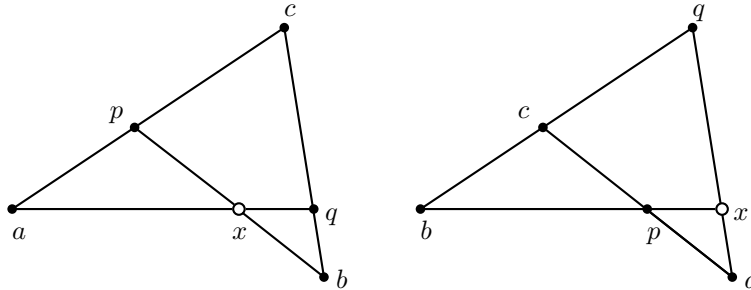


Figure 3: Inner Pasch (left) and outer Pasch (right). Line  $pb$  meets triangle  $acq$  in one side  $ac$ , and meets an extension of side  $cq$ . Then it also meets the third side  $aq$ . The open circles show the points asserted to exist.

to change now.

### Euclid's Postulate 5

Euclid's "parallel postulate", or "Euclid 5", is a postulate rather than an axiom, because it asserts that two lines meet, i.e., there exists a point on both lines. The hypothesis as Euclid stated the postulate involves angles. We use instead a "points-only" version. Then Euclid's version becomes a theorem.

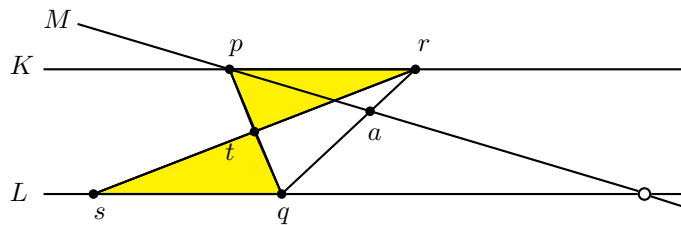


Figure 4: Euclid 5. Transversal  $pq$  of lines  $M$  and  $L$  makes corresponding interior angles less than two right angles, as witnessed by  $a$ . The shaded triangles are assumed congruent. Then  $M$  meets  $L$  as indicated by the open circle.

Most modern geometry textbooks replace Euclid 5 by "Playfair's axiom" (introduced by Playfair in 1729), which asserts the uniqueness of a line parallel to a given line  $AB$ , through a point  $P$  not collinear with

$AB$ . This also becomes a theorem in our development. Although it does not occur as a proposition in Euclid, it is several times used implicitly in Euclid's proofs.

17

## Euclid's Postulate 4

Euclid 4 says "all right angles are equal." The definition of a right angle is this:  $ABC$  is a right angle if there is a point  $D$  such that  $\mathbf{B}(A, B, D)$  and  $AB = DB$ . It has been claimed since the time of Proclus that Euclid 4 is provable, but since the axioms and definitions were not so precise, we are not certain that any of the alleged proofs could be counted as correct until the proof in Tarski's system presented in [35]. In our system this is a difficult proof, depending on the fact that both reflection in a point and reflection in a line are isometries (preserve congruence and betweenness). The proof has to work without a dimension axiom. It is a very beautiful proof and obviously much deeper than the "proofs" given by Proclus and Hilbert. The beautiful part of the proof (after the observation that reflections are isometries) is contained in Satz 10.15 of [35].

Even though this proof is difficult, it would clearly be a flaw to assume Euclid 4 as an axiom, when it can in fact be proved. Therefore we prove it, rather than assume it.<sup>18</sup>

## Axioms involving circles

Euclid had no axioms about circles. There are three axioms in the literature:

*circle-circle*: if circle  $C$  has one point inside circle  $K$  and one point outside, then there is a point on both circles.

*line-circle*: if line  $L$  has a point  $P$  inside circle  $K$ , then there are two points  $A$  and  $B$  on both  $L$  and  $K$ , such that  $P$  is between  $A$  and  $B$ .

*Segment-circle*: if line  $L$  has a point  $A$  inside circle  $K$  and a point  $B$  outside, then there is a point on  $K$  between  $A$  and  $B$ .

The reader should bear in mind that in the absence of any dimension axioms, a "circle" is "really a sphere", or even some kind of "hyper-sphere".

circle-circle is used twice in Book I, once in Prop. I.1 and once again in Prop. I.22, which shows how to construct a triangle with sides congruent to given lines. (The third vertex is the intersection point of two circles

---

<sup>17</sup> Explicitly: Playfair is used directly in propositions 44, 45, 47, and indirectly in 37, 38, 42, 46; and more indirectly in 39, 40, 41, 42, 48; so overall it is used in 39-48 except 43. Euclid 5 is used directly in 29, 39, 42, 44 and indirectly in 29-48 except 31, which is the existence of the parallel line. Euclid should have proved 31 before 29, to emphasize that Euclid 5 is not needed for it.

<sup>18</sup>One may well ask, if we find it necessary to prove Euclid 4 "just because we can", why do we not find it necessary to prove one of the two Pasch axioms, inner and outer Pasch, from the other "just because we can"? The answer is that we still need one of them as an axiom, and the same intuition that justifies one of them also justifies the other. Therefore there is no *conceptual* economy in reducing the number of axioms by one. But proving Euclid 4 does offer a conceptual simplification.

with the specified radii.) Although Euclid does not explicitly mention the axiom, both its applicability and necessity are clear, so we take circle-circle as an axiom.

line-circle is used only once in Book I, in Prop. I.12, the construction of a “dropped perpendicular.” We might also consider “one-point line-circle”, in which the conclusion is weakened to assert only the existence of a single point common to  $L$  and  $K$ . Since this axiom is inadequate for the application to I.12, we do not consider further the idea of using it instead of line-circle.

Segment-circle has been suggested as an axiom by many authors, including Tarski (see [38]). But a detailed study shows that it is inadequate; an irremovable circularity arises in formalizing Euclid without a dimension axiom. If we try to construct dropped perpendiculars (Euclid I.12) using segment-circle continuity, to check the hypotheses we need the triangle inequality (I.20). But I.19 is needed for I.20, and I.7 for I.19. In Prop. I.7, the two triangles that are supposed to coincide might lie in different planes, but for the hypothesis that they lie on the same side of a line, a hypothesis that Euclid stated but never proved. Even so, I.7 is more difficult to prove than Euclid thought, since he took for granted the fact that an angle cannot be less than itself, but that principle is actually the essential content of I.7. Ever since Hilbert [25], angle inequality has been regarded as a defined concept, and proving I.7 then requires dropped perpendiculars (I.12) (or at least, we could not do without I.12). But this is circular. The conclusion is that segment-circle continuity is not a suitable axiom to use in formalizing Euclid’s proofs.<sup>19</sup>

The fact is that each of line-circle and circle-circle implies the other, in the presence of the other axioms of Euclid. In the interest of following Euclid fairly closely, we simply take both as axioms: circle-circle is used in I.1 and I.7, while line-circle is used in I.12, and both those proofs are far simpler than the proofs of line-circle and circle-circle from each other.

Even though we take both as axioms, it is worthwhile to discuss the equivalences. The proofs can be found in [24]; see also the last section of [2]. The proof of line-circle from circle-circle relies on dropped perpendiculars, which in Euclid is I.12, proved from line-circle. Therefore, a proof of line-circle from circle-circle must rely instead on Gupta’s circle-free perpendicular construction [23, 35], carrying us far beyond Euclid. The only known synthetic proof of circle-circle from line-circle uses the “radical axis” [37].

These proofs are long, and can be found, for example, in [24].<sup>20</sup>

---

<sup>19</sup>Line-circle continuity does not suffer from this problem, as the triangle inequality is not required to drop perpendiculars. Of course, as Gupta showed [23], one can construct dropped perpendiculars without mentioning circles at all, so there is no formal result that one circle axiom is better for I.7 than another, as none at all is actually needed. We merely say that Euclid’s proof can be repaired with line-circle, but not with segment-circle.

<sup>20</sup>The argument we gave for proving (rather than assuming) Euclid 4 was that it provides a conceptual simplification to prove it. It does seem that reducing circle-circle to line-circle provides a conceptual simplification, so we really cannot offer a philosophically convincing argument for not computer-checking that reduction. But Euclid’s proofs clearly use both axioms.



## Equal Figures in Euclid

Euclid used the word “figure” to mean what we now call a simple closed polygon. A figure is “contained by” straight lines. In practice Book I needs only triangles and quadrilaterals. Euclid used the word “equal” to denote a relation between figures that he does not define. Nor did Euclid give any explicit axioms about “equal figures”; he treated these as special cases of the common notions, such as “the whole is equal to the sum of the parts”, where the “parts” are figures and the “sum” is the union. Occasionally he uses without explicit mention a few further axioms, such as “halves of equals are equal.”

Book I culminates in the Pythagorean theorem, which Euclid states using the notion of equal figures.<sup>21</sup> Although we formalized only Book I in the work reported here, all the propositions in Book II are about “equal figures”, so a correct formulation of the notion is critical.

Exactly what notion Euclid had in mind is not perfectly clear. One possible interpretation is that equal figures are figures with the same area. That is not a first-order notion, because it involves measuring areas by numbers. Alternately, after introducing “segment arithmetic” geometrically, one can define area geometrically, but that is a very great deviation from the path taken by Euclid.

Another possible interpretation of “equal figures” is the notion of “equal content”, explained on p. 197 of [24], which involves cutting figures into a finite number of pieces and reassembling them. That is also not a first-order notion, because of the “finite number of pieces” part. Hence it is irrelevant for our purposes, and we need not go into the details of the definition.

Conclusion: the definitions of “equal figure” that we find in the literature all suffer from one of the following defects:

- (i) Not being first order, because of requiring the concept of real number.
- (ii) Not being first order, because of requiring the concept of natural number (even just for equality of triangle and quadrilaterals).
- (iii) First order, but requiring the geometrical definition of coordinates and arithmetic (addition and multiplication of finite lines), which goes well beyond Euclid.

The introduction of geometrical arithmetic has already been proof checked in [11], from Tarski’s axioms; and we checked that Tarski’s axioms are equivalent to those used in this paper, so approach (iii) has already been proof checked.

We are, of course, not the first ones to face these difficulties. Hartshorne lists (p. 196 of [24]) the properties of “equal figures” that Euclid’s proofs use. Not all the properties in that list are first order. Our approach to the treatment of “equal figures” is to treat “equal triangles” and “equal figures” (that is, equal quadrilaterals) as primitive relations, and give first-

---

<sup>21</sup>In fact, he says that the squares on the sides together equal the square on the hypotenuse. But what he proves is that the square on the hypotenuse can be divided into two rectangles, each of which is equal the square on one of the sides; so the further notion of two figures together being equal to a third is not really needed.

order axioms for them, sufficient to account for Euclid’s proofs. These axioms are first order versions of Hartshorne’s.

Since we use variables only for points, not for figures, we must use two relations: **ET** for “equal triangles” and **EF** for “equal quadrilaterals” or “equal figures”; it is only for quadrilaterals, but **EQ** is already taken. The axioms for these two relations say that **ET** and **EF** are equivalence relations; that the order of vertices can be cyclically permuted or reversed, preserving equality; that congruent triangles are equal; that if we cut equal triangles off of equal quadrilaterals, producing triangles, the results are equal; or if the cuts produce quadrilaterals, the results are also equal. Then we have “paste” axioms that allow for pasting equal triangles onto equal triangles; if the results are quadrilaterals they are equal, provided also that the triangles do not overlap, which can be ensured by a hypothesis about vertices lying on opposite sides of the paste-line. Similarly, if pasting equal triangles onto equal quadrilaterals produces quadrilaterals, they are equal.

Finally we need an axiom that enables us to prove that certain figures are *not* equal; all the axioms mentioned so far hold if all figures are equal. Such an axiom was introduced by de Zolt (see [24], p. 201). But de Zolt’s formulation is not first order. Instead we take a special case: if  $ABC$  is a triangle, and  $DE$  is a line that cuts the triangle (in the sense that  $\mathbf{B}(A, D, B)$  and  $\mathbf{B}(B, E, C)$ ), then neither of the two resulting pieces  $ADE$  or  $CDE$  is equal to  $ABC$ ). This turns out to be sufficient. A complete list of our axioms, including all the equal-figures axioms, is in the Appendix.

The sudden introduction, halfway through Book I, of a large number of additional axioms, might be considered a point of imperfection; and if it could be improved by giving a first-order, synthetic definition of “equal figures”, and proving the required properties of equal figures without introducing new axioms, that would be desirable. We do have a candidate for such a definition, based on Euclid I.44, but the verification that the axioms hold for that definition seems to be as difficult as the definition and correctness proofs for segment arithmetic. It therefore goes beyond our aim in this paper, which is to give a formal account of Euclid’s definitions, axioms, propositions, and proofs. Since Euclid’s proofs use axioms about equal figures (even if they are listed as axioms), that is what we do too.

## Book Zero and filling in Book I

We proved more than 230 theorems, including the 48 propositions of Book I. To list these theorems in the format used in the Appendix requires 14 pages, and since the files containing these theorems are accessible (as well as the proofs), we elected not to list them all. Still we wish to give the reader some idea of the additional theorems that we had to supply. We use the phrase “Book Zero” informally to encompass those theorems that seem to come before Book I, in the sense of being used in Book I and not depending on Book I themselves. Book Zero begins with properties of congruence and betweenness; several important and often-used lemmas are about the order of four points on a line, when two betweenness relations

are known between them. (There is one axiom about that, and the rest of the relevant propositions can be proved.) There are variations on the 5-line axiom; there are theorems about collinearity and non-collinearity; there is the definition of “less than” for finite lines, and the ordering properties of that relation and how it respects congruence (or equality) of finite lines. Lying on ray  $AB$  (which emanates from  $A$  and passes through  $B$ ) is a defined relation; there are lemmas about how it relates to betweenness and to collinearity. We can “lay off” a finite line along a ray, and the result is unique. Euclid says we can “add equals to equals” as a common notion; but we prove it in a lemma called **sumofparts**. There is also **differenceofparts** and **subtractequals**. Equality and order of angles are defined concepts, and we have to prove their fundamental properties in lemmas such as **ABCequalsCBA**, **equalanglestransitive**, and so on. The “crossbar theorem” is also part of Book Zero. The notions of “same side” and “opposite side” are defined, and their fundamental properties proved, including the plane separation theorem, according to which if  $C$  and  $D$  are on the same side of  $AB$ , and  $D$  and  $E$  are on opposite sides of  $AB$ , then  $C$  and  $E$  are on opposite sides of  $AB$ ; that is, there is a point collinear with  $AB$  that is between  $C$  and  $E$ . This is where we pass out of Book Zero, however, since the proof of that theorem requires constructing a midpoint, which is Proposition 10.

Book Zero comprises about seventy theorems; Euclid’s Book I has 48; we proved an additional hundred of so theorems that are needed to prove Euclid’s 48, or are variants of those propositions, but whose proofs use some of Euclid’s propositions as well. Let us give a typical example: the lemma we call **collinearbetween** is used 19 times in our development, including in Propositions 27,30,32,35,44, and 47. That lemma says that if two lines  $AB$  and  $CD$  are parallel, and there is a point  $E$  between  $A$  and  $D$  that is also collinear with  $BC$ , then that point  $E$  is actually between  $B$  and  $C$ . That is not trivial to prove, and Euclid simply assumes that it is so, because it appears so in the diagram. The names of some other lemmas will be illustrative:

**droppedperpendicularunique**, **angleordertransitive**,  
**angleorderrespectscongruence**, **angletrichotomy**.

Prop. I.12 (dropped perpendicular) has to precede Prop. I.9 (angle bisection), because Euclid’s proof of angle bisection via I.1 cannot be corrected. Once perpendiculars are available, we prove **pointreflectionisometry** and **linereflectionisometry**, and use them to carry out Szmielew’s proof of Euclid’s Postulate 4 (all right angles are equal). Euclid fails to state **legsmallerhypotenuse**, which is needed to prove another fact about right triangles that Euclid uses without proof: the foot of the perpendicular from the right angle to the hypotenuse actually lies between the two endpoints of the hypotenuse. Towards the end of Book I, the steps of the proofs are more cavalier, and the omitted lemmas are more difficult; for example Euclid omitted to state and prove that a square is a parallelogram.<sup>22</sup>

---

<sup>22</sup>A square has four right angles and equal sides. A parallelogram has both pairs of opposite sides parallel.

## 6 Formal representation of Euclid

We wanted to write down our axioms, definitions, postulates, lemmas, and propositions in a form that would be easy to manipulate by computer. We chose to use strings to represent all these things. Euclid used only one-character variable names, and we did the same. In that case there seemed no need for commas and parentheses; in other words we used Polish notation. There were, however, more than 26 relations to consider, so we used in all cases 2-character names for the relations. For example, we write  $\mathbf{B}(a, b, c)$  in the form **BEabc**.  $AB = CD$  is written **EEABCD**. We used **EQAB** to represent  $A = B$ , so we could not use **EQ** a second time, and chose **EE** instead. “Parallel” becomes **PR**, as in **PRABCD**. There are quite a few of these abbreviated two-character names, but that is enough to convey the idea. The point is that every formula is a string. Conjunctions and disjunctions begin with **AN** or **OR** and the subformulas are separated by **+** or **|**, respectively. Negations are formed with **NO**.

Then we define classes **Axiom**, **Definition**, and **Theorem**, each of which has fields called **label**, **hypothesis**, **conclusion**, and **existential**. The **label** field is used for the name. The **hypothesis** and **conclusion** fields each contain an array of formulas, or a single formula. The **existential** field contains an empty string by default, and if it isn’t empty, that means that it contains a list of variables that are supposed to be existentially quantified in the conclusion.

We have thus defined a subset of first-order predicate logic. Specifically, our formulas have no function symbols, and only existential quantifiers; universal quantification over the free variables is left implicit. Nested quantifiers do not occur. Polish notation, one-character variables, and two-character predicate names make it easy to manipulate these formulas as strings and arrays of strings, and substitutions can be coded as arrays, making unification possible by regular-expression matching. Every modern programming language has useful libraries for this sort of thing.

## 7 Formal proofs

Each proof is a list of lines. Each line contains a formula and optionally a justification. The proof is kept in a **.prf** file whose name gives the label of the theorem it is intended to prove. The proof begins with a sequence of unjustified lines that must repeat the hypotheses of the theorem. It ends with a line that is the conjunction of the conclusions of the theorem, or the sole conclusion if there is only one. The first line after the hypotheses must have a justification. Any unjustified lines in the rest of the proof must either be repetitions of earlier lines or must follow by logic alone from some earlier lines. Justifications follow the pattern **kind:label**, where **kind** is either **defn**, **axiom**, **postulate**, **proposition**, or **lemma**, and **label** is the label of an item of the specified kind. An axiom, postulate, or definition can be used anywhere, but a lemma or proposition cannot be used anywhere, because circular arguments must be prevented.

We avoid circularity by having a “master list” of lemmas and propositions, which are to be proved in the specified order. A valid proof (of a

certain item in the master list) is only allowed to reference previous items in its justified steps.

Euclid's proofs, and ours, make use of arguments by contradiction and cases. We will now explain the syntax we used. An argument by contradiction is introduced by a line with the justification **assumption**. After some steps of proof, this line must be matched by a line with the justification **reductio**. This line must contradict the assumption line. By saying that  $A$  and  $B$  contradict each other, we mean that one of them is the negation of the other. We found it helpful to indent the lines between the assumption and the reductio labels, especially when nested arguments by cases or reductio occur.

The syntax for cases is as illustrated in Fig. 5. Some of the cases also illustrate proof by contradiction. The proof fragment begins with an unjustified **COEFH**, which means  $E$ ,  $F$  and  $H$  are collinear. This is repeated from earlier in the proof. The second line is a disjunction, giving the definition of collinear. The argument by cases begins on the third line. The cases are exhaustive, by the second line. Each case begins with a case number and closes with the keyword **qedcase**. Indentation is for human use only.

In the proof, the only symbol that has not yet been explained is **NE**. This means “not equal”, and has the same meaning as **NOEQ**. Similarly, **NC** for “noncollinear” has the same meaning as **NOCO**.

We wrote more than two hundred formal proofs in this syntax. For convenience, we introduce a name for this subset of first-order logic: **Euc**, the first three letters of Euclid.

```

COEFH
OREQEF|EQEH|EQFH|BEFEH|BEEFH|BEEHF defn:collinear
cases BEEHF:EQEF|EQEH|EQFH|BEFEH|BEEFH|BEEHF
case 1:EQEF
  NOBEEHF assumption
  NEEF
  BEEHF reductio
qedcase
case 2:EQEH
  NOBEEHF assumption
  NEEH lemma:inequalitysymmetric
  BEEHF reductio
qedcase
case 3:EQFH
  NOBEEHF assumption
  NEFH lemma:inequalitysymmetric
  BEEHF reductio
qedcase
case 4:BEFEH
  NOBEEHF assumption
  NOBEFEH
  BEEHF reductio
qedcase
case 5:BEEFH
  NOBEEHF assumption
  NOBEEFH
  BEEHF reductio
qedcase
case 6:BEEHF
qedcase
BEEHF cases

```

Figure 5: An illustrative proof fragment, showing the syntax of formulas and proofs.

## 8 Checking the proofs by computer

The proofs described in this paper only need a rather weak logic. There are no function symbols, and all the statements have a very restricted form. That made it easy to write a custom-built proof checker, or “proof debugger”, that we used while developing the formalization. That tool also checks that we stay within the bounds of that logic.

HOL and Coq use an architecture that guarantees a much higher reliability.

This is called the *LCF architecture*, after the LCF system from the seventies that pioneered the approach. This architecture divides the system in a small kernel (or *logical core*) and the rest of the code. By the use of abstract datatypes, the correctness of the mathematics is then guaranteed by the correctness of the kernel. Whatever errors the rest of the code of the system may contain, the statements claimed to have been proved will indeed have been proved. In the case of the HOL Light system (one of the incarnations of HOL), the correctness of this kernel has even been *formally* proved (using the HOL4 system, another incarnation of HOL), which gives an *extremely* high guarantee that the system will not have any logical errors.

Our procedure for proof-checking Euclid was thus as follows:

- Write formal proofs in the **Euc** language, simultaneously checking and debugging them with our custom proof debugger.
- Translate these proofs into HOL Light or Coq syntax by means of simple scripts.
- Check the resulting proofs in HOL Light and Coq

The devil might ask whether we have lost something by using higher order logic to check first order proofs. While it may appear so at first glance, actually higher order logic itself ensures that we have not. Consider: in higher order logic (both in Coq and HOL Light) we proved that for any type of points, and couple of Betweenness and Congruence predicates verifying Tarski’s axioms we can define the predicates of **Euc** and prove the axioms. In these statements there is a second-order quantification over the predicates used to interpret betweenness and congruence. In essence we have proved that the axioms hold in any model. Then by Gödel’s completeness theorem, the theorems are actually first order theorems. However, it is a general feature of higher order theorem provers that they do not directly check first-order proofs. Moreover, we used the Leibniz definition of equality, so indeed our translated proofs are *not* first order. However, we did check the first order proofs directly in our custom proof checker before translating them to higher order logic.

Our debugger also counted the number of inferences. Proofs of more than 200 inferences were not uncommon, but the majority were under 100 inferences.

We also wrote code that analyzed the dependencies between lines of a given proof. This enabled us to identify and eliminate lines that were never subsequently used. We follow Euclid in sometimes repeating previously deduced lines just before applying a proposition, to make it apparent

that the required instances of the hypotheses have indeed been derived. These lines, of course, are technically eliminable; but mainly we wanted to eliminate “red herring” lines that were actually irrelevant. The automatic detection of such lines was useful.

## 9 Checking the proofs in HOL Light

To ensure the correctness of the Euclid formalization from the language **Euc** to HOL Light, we built a very small custom checker on top of HOL Light, and used that to check our work for correctness in HOL as well. The source of this proof checker, `proofs.ml`, has about a hundred lines, which then is used to check a translation of the formalization into syntax that HOL Light can process of almost twenty thousand lines, `michael.ml`. This last file is created from the original proof files described above by two small *ad hoc* scripts, a PHP script called `FreekFiles.php` and a Perl script called `FreekFiles.pl`.

In the HOL system, all input (even the proofs) always consists of executable ML source code. The proofs of each lemma in our case is checked by calling a function `run_proof` (implemented in `proofs.ml`) on the statement of the lemma and a list items of a custom datatype called `proofstep`. The ML definition of this datatype is shown in Fig. 6. The output of this function is a HOL `thm`, a proved statement. In other words, the ML type of the function that is used here is:

```
run_proof : term -> proofstep list -> thm.
```

The list of `proofsteps` corresponding to the proof fragment from Fig. 5 is shown in Fig. 7.

In that example some of the choices on how to translate the statements from the proof to the HOL logic can be seen. For instance, we had to decide whether to translate `EQ` to the standard HOL equality, or to have it be a custom relation. We chose to make use of the features of standard first order logic with equality, but nothing beyond that. That means that we translated equality to the built-in equality of the logic, and translated

```
type proofstep =
  | Known of term
  | Step of term * thm
  | Equalitysub of term
  | Assumption of term
  | Reductio of term
  | Cases of term * term
  | Case of int * term
  | Cases_ of term
  | Qedcase;;
```

Figure 6: The ML datatype used when checking the proofs with HOL.



```

Known 'CO E F_ H';
Step ('E:point = F_ \\/ E:point = H \\/ F_:point = H \\/ BE F_ E H \\/ BE E F_ H \\/ BE E H F_',
Cases ('BE E H F_', 'E:point = F_ \\/ E:point = H \\/ F_:point = H \\/ BE F_ E H \\/ BE E F_ H \\/ BE E H F_');
Case (1, 'E:point = F_');
Assumption '~(BE E H F_);
Known '~(E:point = F_);
Reductio 'BE E H F_';
Qedcase;
Case (2, 'E:point = H');
Assumption '~(BE E H F_);
Step ('~(E:point = H)', lemma_inequalitysymmetric);
Reductio 'BE E H F_';
Qedcase;
Case (3, 'F_:point = H');
Assumption '~(BE E H F_);
Step ('~(F_:point = H)', lemma_inequalitysymmetric);
Reductio 'BE E H F_';
Qedcase;
Case (4, 'BE F_ E H');
Assumption '~(BE E H F_);
Known '~(BE F_ E H);
Reductio 'BE E H F_';
Qedcase;
Case (5, 'BE E F_ H');
Assumption '~(BE E H F_);
Known '~(BE E F_ H);
Reductio 'BE E H F_';
Qedcase;
Case (6, 'BE E H F_');
Qedcase;
Cases_ 'BE E H F_';

```

Figure 7: The HOL counterpart of the proof fragment in Fig. 5.

NE identical to NOEQ and NC identical to NOCO. Therefore the translation does not have predicates EQ, NE and NC. As a small optimization, NONC was translated without a double negation.

Within HOL we used an axiomatic approach. That is, we added our axioms to the HOL axioms. That way, we will be verifying that the theorems of Euclid follow from those axioms, rather than (for example) that they are true in  $\mathbb{R}^2$  or  $\mathbb{R}^n$ . This raised the number of HOL axioms by 40, from the original 3 to 43. Before stating these 40 axioms we also added two new primitive types. `point` and `circle`, and five new primitive predicates, BE, EE, CI, ET and EF.

There are some differences between what was taken to be axioms in the original version of the formalization, and what are axioms in the HOL version. The definitions of the predicates were originally axioms, but in the HOL version are actual HOL definitions, with the ‘axioms’ being the equivalences that these definitions produce. Two exceptions for this approach are `defn:unequal` and `defn:circle` of which the first is omitted (it is not used anywhere) and the second is an axiom (as it does not have the shape of a HOL definition). Also, `defn:inside`, `defn:outside` and `defn:on` are still axioms, because these also state that the defining property does not depend on the points that give the circle, which one does not get from just a definition.

Of the common notions, `cn:equalitytransitive`, `cn:equalityreflexive`, `cn:stability` and `cn:equalitysub` are not axioms, but proved statements, because they only involve equality and are part of the logic.

The LCF architecture of HOL Light only guarantees that the proofs are valid in the higher order logic of HOL Light. However, the implementation in `proofs.ml` only uses first order tactics (most notably `MESON`, which is the main tool used for checking the steps), which shows that the proofs are actually first order. Whether the proofs are correct with respect to the ‘natural’ interpretation of the proof steps (whatever that is) is not checked in the HOL version.

The full check of the twenty thousand lines is uneventful, but not very fast, mostly because of the use of the rather heavy `MESON`. It takes several minutes. At some points `MESON` has to work very hard (it loses the fact that the conclusion of the lemma exactly matches the step being proved, and is trying many possible ways to unify the parts). For these cases `proofs.ml` contains two custom lower level tactics, `SUBGOAL_UNFOLD_TAC` and `SUBGOAL_MATCH_TAC`, which do not use `MESON`. Potentially the use of `MESON` could be avoided altogether, but then the code in `proofs.ml` would be much more involved, with backtracking to find the relevant previous proof steps. Still this might be useful, as then it would establish which of the previous proof steps are actually used when checking a step. This would make a practical translation to a Mizar version of this formalization possible.

```

Lemma proposition_41 :
  forall A B C D E, PG A B C D -> Col A D E -> ET A B C E B C.
Proof.
intros.
assert (ET A B C E B C).
by cases on (eq A E /\ neq A E).
{
  assert (ET A B C A B C) by (conclude axiom_ETreflexive).
  assert (ET A B C E B C) by (conclude cn_equalitysub).
  close.
}
{
  assert (Par A B C D) by (conclude_def PG ).
  assert (Par A D B C) by (conclude_def PG ).
  assert (Col D A E) by (forward_using lemma_collinearorder).
  assert (Par B C A D) by (conclude lemma_parallelsymmetric).
  assert (Par B C D A) by (forward_using lemma_parallelflip).
  assert (neq E A) by (conclude lemma_inequalitysymmetric).
  assert (Par B C E A) by (conclude lemma_collinearparallel).
  assert (Par B C A E) by (forward_using lemma_parallelflip).
  assert (Par A E B C) by (conclude lemma_parallelsymmetric).
  assert (ET A B C E B C) by (conclude proposition_37).
  close.
}
close.
Qed.

```

Figure 8: Example proof in Coq

## 10 Checking the proofs in Coq

### 10.1 Formalizing the Axioms

The axioms in our axiomatization of Euclid are of two kinds: axioms related to definitions and others. Axioms that serve as definition as those of the form:  $\forall x, P(x) \iff Q(x)$ . We translate them to a proper Coq definition to reduce the number of axioms. Technically, in Coq, we did not use the `Axiom` keyword. Because axioms are similar to global variables in a programming language, they reduce the re-usability of the code. We sorted the axioms into several groups, and defined them in Coq using type classes. Then, the axioms are given as so-called section variables of Coq, a mechanism which allows to have the axioms as an implicit assumption for each lemma. For propositions I.1 to I.28 and I.30 we do not use the fifth postulate of Euclid. For propositions I.1 to I.34, we do not need the equal-figure axioms. Avoiding the `Axiom` keyword allows us to reuse the proofs in a different setting by proving the axioms in Coq as a second-order property either from another axiom system or by constructing a model (see Sec.11.1).

**Equality** We model the equality using Coq’s built-in equality: Leibniz’s equality. We could also have assumed an equivalence relation and substitution properties for each of the predicates of the language.

### 10.2 Verifying the proofs

Proof assistants differ in their mathematical foundations (e.g. type theory, higher order logic (HOL), or set theory) and their proof language. In procedural style proof assistants (e.g. Coq and HOL Light), proofs are described as a sequences of commands that modify the proof state, whereas in proof assistants that use a declarative language (e.g. Mizar and Isabelle), the proofs are structured and contain the intermediate assertions that were given by the user and justified by the system.

We wrote a script to translate the proofs to Coq’s language. The translation is relatively easy as the proofs steps used by the proof debugger are small.

Our translation generate a proof in the traditional language of Coq, not in the declarative language introduced by Corbineau [16] because this language is not maintained. But, the formal proofs we generate are in the declarative style: for case distinctions we give explicitly the statement which is used instead of the name of the hypothesis, the proofs are purely in the forward chaining style, based on sequences of applications of the standard Coq tactic `assert`. We do not introduce hypothesis numbers. The existential statements and conjunctions are eliminated as soon as they appear by introducing the witness and decomposing the conjunction. The assertions are justified using some ad-hoc tactics which use some automation and congruence closure. We had to circumvent some weakness of Coq’s automation. Coq is not able to use efficiently lemmas of the form  $\forall xyz, Pxyz \rightarrow Pyzx \wedge Pxyz$ , because the `apply` tactic will always choose to unify the goal with the first term of the conjunction. The standard

way to state such a lemma in Coq is to split the lemmas in two parts. But, we did not want to modify the original formalization of the lemmas. Hence, to circumvent this limitation, we verified these proof steps using a tactic based on forward-chaining. The case distinction tactic, allows to distinguish cases on previously proved disjunctions or disjunctions which are classical tautologies. The tactic can deal with n-ary disjunctions. The proof script is structured using curly brackets and indentation. Each proof step of the original proof corresponds to one proof step in Coq, except that steps for which there is no justification in the original proofs. Those steps corresponds the natural deduction rule for introduction of implication and are implemented using the standard Coq tactic `intro`. Figure 8 provides an example of a proof in Coq’s language enriched by the tactics to ease the verification. Some of the predicates have been renamed into longer names to enhance readability and to match the names used in the GeoCoq library. The proofs are verified using classical logic. The proofs takes approximately 45 minutes to compile on a laptop.

### 10.3 Verifying the statements

When evaluating a formalization, even if we trust the proof checker, we need to check that statements are formalized faithfully. Usually the only method we can use for this process is to check the statements by human inspection and trust the reviewer to also check the statements. For this formalization, we were lucky, as many statements had been formalized independently by the first author and the GeoCoq team. To improve the confidence in the formalization, we compared the two formalizations of the statements to detect potential defects. We detected only minor differences in some of the statements.

## 11 The axioms hold in $\mathbb{R}^2$

The axioms fall into two groups: those that are variants of Tarski’s axioms A1–A10, and the equal-figure axioms. In order to make sure that there is no mistake in the axiomatization, we wished to check formally that the axioms hold in the Cartesian plane  $\mathbb{R}^2$ . For reasons of convenience, we checked that for the first group of axioms in Coq, and for the equal-figure axioms in HOL Light. We will discuss these two verifications separately.

It has already been checked [7] that Tarski’s axioms hold in  $\mathbb{R}^2$ . In Tarski’s A1–A10, without dimension axioms, one can (formally) verify our axioms, which are mainly different from Tarski’s by using strict betweenness and hence avoiding degenerate cases. This has also been verified in Coq.

### 11.1 Verifying the Tarski-style axioms in Coq

Since there is no dimension axiom, the “intended model” is  $\mathbb{R}^n$ , for any integer  $n > 1$ . More generally, we wish to prove that if  $\mathbb{F}$  is a Euclidean field, then  $\mathbb{F}^n$  satisfies our axioms. We break this claim into four parts.

(i) our axioms for neutral geometry can be derived from the corresponding Tarski’s axioms.

(ii) circle-circle, circle-line, and Euclid 5 can be derived from the corresponding axioms from Tarski.

(iii) Tarski’s axioms hold in  $\mathbb{F}^n$ .

(iv) the equal-figure axioms hold in  $\mathbb{F}^n$ .

The reason for passing through Tarski’s axioms is that it has already been shown that Tarski’s axioms for neutral geometry hold in  $\mathbb{F}^2$ , when  $\mathbb{F}$  is a Pythagorean ordered field.<sup>23</sup> Specifically, Boutry and Cohen have formalized the proof that the Cartesian plane over a Pythagorean ordered field is a model of our formalization of Tarski’s axioms A1–A10 using `ssreflect` [5]. The model has been extended to verify the circle-circle intersection axiom, by assuming that the field is Euclidean. Boutry and Cohen have since extended their work to prove that Tarski’s axioms, except dimension, hold in  $\mathbb{F}^n$ .

Ad (ii). We chose to assume the circle-circle intersection property and use the formal proof obtained by Gries and the second author that the circle-line intersection property can be derived from circle-circle intersection, even without assuming a parallel postulate [22]. Gries also formalized the proof that circle-circle intersection can be derived from the continuity axiom of Tarski (Dedekind cuts restricted to first-order definable sets). As **Euc** axiom system assumes that we have a sort for circles, we need to define the type of circles from Tarski’s axioms (and Coq’s Calculus of Inductive Constructions CIC). The type of circles is defined as the triple of points (`*` is interpreted in this context as the Cartesian product).

**Definition** `Tcircle` : Type := Tpoint\*Tpoint\*Tpoint %type.

Then the predicate `CI` can be defined by:

**Definition** `CI` (J:Tcircle) A C D := J=(A,C,D).

Then the relation expressing that a point is on a circle can be defined by destructing the triple:

**Definition** `OnCirc` P (C:Tcircle) :=  
`match C with`  
`(X,A,B) => tarski_axioms.Cong X P A B`  
`end.`

For Euclid 5, we rely on the proofs of equivalence between different versions of the parallel postulates studied previously by the first author [3] and formalized in Coq, by Boutry, Gries, Schreck and the second author [8].

## 11.2 Verifying the equal-figure axioms

We here outline the verification that the the equal-figure axioms hold in  $\mathbb{R}^2$ . We did not verify them in  $\mathbb{F}^n$  or even  $\mathbb{R}^n$ , because we wanted to use the (scalar) cross product in  $\mathbb{R}^2$  and the existing tools for real algebra in HOL Light.

---

<sup>23</sup>Recall that a Pythagorean field, is a field where some of squares are squares.

We interpret points as members of the Cartesian product  $\mathbb{R} \times \mathbb{R}$ . We write that here as  $R^2$ , although in HOL Light the notation  $R^2$  does not mean the Cartesian product. Then we define the dot product and the two-dimensional cross product as usual:

$$\begin{aligned}(a, b) \cdot (c, d) &= ac + bd \\ (a, b) \times (c, d) &= ad - bc\end{aligned}$$

Twice the signed area of a triangle  $abc$  is defined by

$$tarea(a, b, c) = (c - a) \times (b - a)$$

Collinearity  $L(a, b, c)$  and betweenness  $B(a, b, c)$  are defined thus:

$$\begin{aligned}L(a, b, c) &\leftrightarrow \text{tarea}(a, b, c) = 0 \\ B(a, b, c) &\leftrightarrow L(a, b, c) \wedge (b - a) \cdot (c - a) > 0 \wedge (b - c) \cdot (a - c) > 0\end{aligned}$$

Twice the signed area of a quadrilateral  $abcd$  is given by the cross product of its diagonals:

$$sarea4(a, b, c, d) = (c - a) \times (b - d)$$

Now we can define  $ET$  (equal triangles) by saying that two triangles are equal if the squares of their signed areas are equal, and similarly we define  $EF$  (equal quadrilaterals) by saying two quadrilaterals are equal if the squares of their signed areas are equal. The area of a triangle or quadrilateral is defined to be the absolute value of the signed area.

These are all the definitions needed to interpret the equal-figures axioms in  $R^2$ . We executed this translation by hand, producing a list of seventeen goals to prove in HOL Light. Eight of these were extremely easy to prove; the others were `halvesofequals`, the two De Zolt axioms, and the cut-and-paste axioms. The remaining axioms are consequences of the additivity of area, so we proved them by first proving the additivity of area, and then deriving the axioms from the additivity of area.<sup>24</sup>

## 12 Previous work on computer checking geometry

Work on computerizing Euclidean geometry began in 1959, in the first decade of the computer age, with the pioneering work of Gelernter [19, 20]. (The reference has a later date because it is a reprinting in a collection.) The axiom system used by Gelernter was not given explicitly, but from the example proofs given, it can be seen that it was, at least in effect, a points-only system. It was a strong axiom system, including for example

---

<sup>24</sup>Full disclosure: as of October 1, 2017, we have not yet completed these proofs. We plan to complete them and remove this footnote before publication.

all the triangle congruence theorems, Euclid 4, some strong but unspecified betweenness axioms. Tarski is not referenced; that is not surprising as Tarski's first publication of his axiom systems was also in 1959. Gelertner's system was claimed to be as good as "all but the best" high school geometry students. Considering the primitive hardware and software of 1959, it was an amazing program. However, its authors stated that they viewed geometry as just one area in which to study heuristic reasoning, and neither the program nor its underlying formal theory ever raised its head above water again.

As far as we know, nineteen years passed before the next work in computerizing geometry; it was 1978 when Wen-Tsün Wu [40] began a series of papers on the subject, culminating in his 1984 book, published only in Chinese, and not available in English until 1994 [41]. He was soon assisted by S. C. Chou [12, 13]. Using coordinates, one reduces a geometry theorem to an implication between polynomial equations. One can then demonstrate the truth of a geometry theorem by algebraic methods. Such algebraic methods, while they may succeed in establishing the validity of a theorem, do not provide a proof from geometric axioms. One might check the correctness of Euclid's *results* this way, but not his *proofs*. One problem with this approach is that it works only for problems that involve equality, not for problems that involve betweenness or inequality. A second problem is that Wu's method requires polynomials with hundreds of variables, and hence does not produce human-comprehensible proofs. Nevertheless, Wu and Wang, Gao, Chou, Ko, and Hussain proved many theorems. Within a few years three different groups began to use Gröbner bases to do the algebraic work, instead of the Wu-Ritt algorithm. See [26], [27], [15].

Sometime after 1984, Chou invented the *area method*, which still uses polynomial computations, but based on certain geometric invariants. The area method, as an algorithm for solving geometry problems, can be used by humans and has even been used to train students for Olympiad-style competitions. [14], p. xi. For the state of the art in area-method implementation as of 1994, see [14], where more than four hundred computer-produced and human-readable proofs are given. However, the steps of these proofs are equations, whose truth is verified by symbolic computation, not by logic. Also, as with Wu's method, inequalities and betweenness cannot be treated.

In the 1980s, Larry Wos experimented with proving geometry theorems in Tarski's theory using the theorem-prover OTTER. Then Art Quaife took up that same project, publishing a paper in 1989 [34], and devoting a chapter of his 1992 book to it. Wos and Quaife left a number of "challenge problems" unsolved by OTTER. In this same decade, the book [35] was published, containing the results from Gupta's thesis [23] and Szmielw's Berkeley course, together constituting a systematic development of "absolute geometry" (no circle or continuity axioms) from Tarski's axioms. This book was quite rigorous, but not (yet) computer-checked. It also did not reach even to the beginning of Euclid. Twenty years later, the first author and Wos returned to this project, and used OTTER to find proofs of all the challenge problems of Quaife, and the first ten chapters of [35] (Part I). But since [35] spends a lot of effort de-



veloping “elementary” results from minimal axioms (no circle axioms and postponing the parallel postulate as long as possible), the propositions of Euclid are not reached. In this project, the more difficult proofs were not found automatically, but instead the theorem prover was used almost like a proof checker, by means of supplying “hints.” Therefore, when we wanted to proceed to proof check Euclid, it seemed appropriate to switch from a theorem prover to a proof checker, which is what we did for this work.

In 2009, Avigad, Dean, and Mumma [1] reported on a formal system for Euclid’s *Elements*. This system is six-sorted (points, lines, circles, segments, angles, and “areas” (figures), and therefore has also a large number of primitive relations, including “same side” and “opposite side”. The axiom system of E differs from ours, because it contains more axioms. They assume 20 constructions axioms, 34 axioms about the two-side, inside and betweenness relations which they call “diagrammatic inferences.” Their system is intended for two-dimensional geometry only, a restriction deliberately avoided in our system. They also assume that distances and areas can be *measured* using a linearly ordered abelian group. We follow Euclid in not assigning measures to distance or area; one may compare distances or areas (figures), but not measure them.

Our aim in this section is to discuss past work on computerization of Euclid, rather than logical formal systems, so our focus is on implementation of the formal system. The intention of the authors was to build an interactive proof checker (with application to education). That proof checker should check “diagrammatic inferences”, allowing the user to make in one step an inference that might require many formal steps in first-order logic. In §6 of the paper, Avigad, Dean, and Mumma say “our analysis should make it possible to design a computational proof checker” based on their system, and “preliminary studies suggest that general purpose tools in automated reasoning are sufficient for the task.” Moreover, a couple of examples of checked proofs are given. In 2011, Ben Northrop implemented an “E proof checker” in Java, which can be found on the web. But there is no archive of checked proofs, or any claim that Book I or any specific body of geometry has been checked. The aims of Avigad, Dean, and Mumma were not along those lines: they wanted to analyze the nature of Euclid’s proofs and account for the role of the diagram in those proofs; computer-checking the correctness was not their primary aim.

Starting in 2007, and still continuing as this is written in 2017, the second author of this paper and Gabriel Braun have been busy computer-checking geometrical theorems in the proof assistant Coq. They verified Pappus’s theorem [11] (which is important for the geometrical definition of arithmetic). They verified that Hilbert’s axioms follow from Tarski’s [10]. With Pierre Boutry, they verified that Tarski’s axioms follow from Hilbert’s [9], and completed [6] the verification of the theorems in Szmielew’s part of [35], which the second author began (with other co-authors) in [36]. Work is currently being done toward checking Euclid’s propositions from Hilbert/Tarski axioms within Coq. This work differs from the work presented in this paper, because the goal is not to verify Euclid’s *proofs* but Euclid’s *statements* using an axiom system as minimal as possible.

For further information, please see the forthcoming survey article [29].

## 13 Conclusion

Our aim was to remove every flaw from Euclid’s axioms, definitions, postulates, and common notions, and give formal proofs of all the propositions in Book I. Did we achieve that aim?

The statements of the postulates and the definitions needed little if any change; it is the axioms and proofs that needed corrections. We replaced Euclid’s axioms and postulates by similar ones in a language similar to Tarski’s, but using strict betweenness. We added line-circle and circle-circle axioms, and both inner and outer Pasch; we added the five-line axiom to enable a correct proof of the SAS congruence criterion (Prop. I.2). We dropped Postulate 4 (all right angles are equal) because it can be proved, and formulated Postulate 5 (the parallel postulate) in our points-only language. Inequality of lines and angles, and equality of angles, become defined concepts and the common notions concerning those concepts become theorems. We used Tarski’s definition of “same side”, an essential concept which Euclid mentioned but neither defined nor considered as a common notion. We think that this choice of axioms is very close to Euclid’s.

With this choice of axioms, we were able to prove Euclid’s propositions I.1 to I.48, except three, which we omitted for good reasons.<sup>25</sup> These proofs follow Euclid as closely as possible, and have been checked in two well-known and respected proof checkers. We have therefore shown beyond a shadow of a doubt that these proofs are correct. Then we checked, again using those same proof checkers, that the axioms we used hold in every plane over a Euclidean field; indeed, since Euclid has no dimension axiom, over every finite-dimensional space over a Euclidean field. In particular all of Euclid’s propositions in Book I *and corrected proofs of those propositions, close to Euclid’s ideas*, are valid, without a shadow of a doubt, in Euclidean two-space and three-space. While many paper-and-pencil formalizations of Euclid have been put forward in the past, we are the first to be able to make this claim. That this was not a superfluous exercise is shown by the many difficulties we encountered, and the fact that we had to prove the propositions in quite a different order than Euclid, and in some cases by different proofs. In this paper we have focused on the axioms and the proof checking. A subsequent publication will present the geometrical difficulties and the repaired proofs.

To play the devil’s advocate, what argument could be made that we did not achieve the aim stated above? One might complain that we proved the propositions in a different order than Euclid did. We had to do that, because we could not prove them in the original order using our axioms. The devil might argue that we should have strengthened the circle-circle axiom to provide for an intersection point of the two circles on a given side of the line joining the centers. With this stronger axiom we could have fixed Euclid’s proof of I.6 (angle bisection) and used it as Euclid did

---

<sup>25</sup>That would be I.2, which is subsumed by our extension axiom, and I.13 and I.17, which require formalizing a theory of angles more than  $180^\circ$ , and are never used.

to bisect a line. But this would amount to assuming, rather than proving, the existence of erected perpendiculars, midpoints, and angle bisectors. Besides, it would not have fixed the other problems we had with the ordering of theorems, and the proof we gave, using midpoints to bisect angles rather than bisection of angles to construct midpoints, is beautiful, even if it was discovered by Gupta more than two thousand years after Euclid.

We think that the devil would be wrong to say we should have strengthened circle-circle, and we therefore claim that we did indeed remove every flaw from Euclid's axioms, definitions, postulates, and common notions, and give correct proofs of the propositions in Book I.

## Appendix

### Common Notions

equalitytransitive  
 hypotheses: EQAC EQBC  
 conclusion: EQAB

congruencetransitive  
 hypotheses: EEPQBC EEPQDE  
 conclusion: EEBCDE

equalityreflexive  
 hypotheses: none  
 conclusion: EQAA

congruencereflexive  
 hypotheses: none  
 conclusion: EEABAB

equalityreverse  
 hypotheses: none  
 conclusion: EEABBA

stability  
 hypotheses: NONEAB  
 conclusion: EQAB

equalitysub  
 hypotheses: EQDA BEABC  
 conclusion: BEDBC

### Definitions

unequal *A and B are distinct points*  
 The definition of NEAB  
 is: NOEQAB

collinear *A, B, and C are collinear*  
 The definition of COABC  
 is: OREQAB|EQAC|EQBC|BEBAC|BEABC|BEACB

noncollinear *A, B, and C are not collinear*  
 The definition of NCABC  
 is: NEAB NEAC NEBC NOBEABC NOBEACB NOBEBAC

circle *X is the circle with center C and radius AB*  
The definition of NEAB  
is: For some X, CIXCAB  
inside *P is inside the circle J of center C and radius AB*  
The definition of CIJCAB ICPJ  
is: For some XY, CIJCAB BEXCY EECYAB EECXAB BEXPY  
outside *P is outside the circle J of center C and radius AB*  
The definition of CIJCAB OCPJ  
is: For some X, CIJCAB BECXP EECXAB  
on *P is on the circle J of center C and radius AB*  
The definition of CIJACD ONBJ  
is: CIJACD EEABCD  
equilateral *ABC is equilateral*  
The definition of ELABC  
is: EEABBC EEBCCA  
triangle *ABC is a triangle*  
The definition of TRABC  
is: NCABC  
ray *C lies on ray AB*  
The definition of RAABC  
is: For some X, BEXAC BEXAB  
lessthan *AB is less than CD*  
The definition of LTABCD  
is: For some X, BECXD EECXAB  
midpoint *B is the midpoint of AC*  
The definition of MIABC  
is: BEABC EEABBC  
equalangles *Angle ABC is equal to angle abc*  
The definition of EAABCabc  
is: For some UVuv, RABAU RABCV RAbau RAbcv EEBUbu EEBVbv EEUvuv NCABC  
supplement *DBF is a supplement of ABC*  
The definition of SUABCDF  
is: RABCD BEABF  
rightangle *ABC is a right angle*  
The definition of RRABC  
is: For some X, BEABX EEABXB EEACXC NEBC  
perpat *PQ is perpendicular to AB at C and NCABP*  
The definition of PAPQABC  
is: For some X, COPQC COABC COABX RRXCP  
perpendicular *PQ is perpendicular to AB*  
The definition of PEPQAB  
is: For some X, PAPQABX  
interior *P is in the interior of angle ABC*  
The definition of IAABCP  
is: For some XY, RABAX RABCY BEXPY  
oppositeside *P and Q are on opposite sides of AB*  
The definition of OSPABQ  
is: For some X, BEPXQ COABX NCABP  
sameside *P and Q are on the same side of AB*  
The definition of SSPQAB

is: For some XUV, COABU COABV BEPUX BEQVX NCABP NCABQ  
 isosceles *ABC is isosceles with base BC*  
 The definition of ISABC  
 is: TRABC EEABAC  
 cut *AB cuts CD in E*  
 The definition of CUABCDE  
 is: BEAEB BECED NCABC NCABD  
 trianglecongruence *Triangle ABC is congruent to abc*  
 The definition of TCABCabc  
 is: EEABab EEBCbc EEACac TRABC  
 anglelessthan *Angle ABC is less than angle DEF*  
 The definition of AOABCDEF  
 is: For some UXV, BEUXV RAEDU RAEFV EAABCDEX  
 togethergreater *AB and CD are together greater than EF*  
 The definition of TGABCDEF  
 is: For some X, BEABX EEBXCD LTEFAX  
 togetherfour *AB,CD are together greater than EF,GH*  
 The definition of TTABCDEFHG  
 is: For some X, BEEFX EEFXGH TGABCDEX  
 tworightangles *ABC and DEF make together two right angles*  
 The definition of RTABCDEF  
 is: For some XYZUV, SUXYUVZ EAABCXYU EADEFVYZ  
 meet *AB meets CD*  
 The definition of MEABCD  
 is: For some X, NEAB NECD COABX COCDX  
 cross *AB crosses CD*  
 The definition of CRABCD  
 is: For some X, BEAXB BECXD  
 tarski\_parallel *AB and CD are Tarski parallel*  
 The definition of TPABCD  
 is: NEAB NECD NOMEABCD SSCDAB  
 parallel *AB and CD are parallel*  
 The definition of PRABCD  
 is: For some UVuvX, NEAB NECD COABU COABV NEUV COCDu COCDv NEuv  
 NOMEABCD BEUXv BEuXV  
 anglesum *ABC and DEF are together equal to PQR*  
 The definition of ASABCDEFpqr  
 is: For some X, EAABCPQX EADEFXQR BEPXR  
 parallelogram *ABCD is a parallelogram*  
 The definition of PGABCD  
 is: PRABCD PRADBC  
 square *ABCD is a square*  
 The definition of SQABCD  
 is: EEABCD EEABBC EEABDA RRDAB RRABC RRBCD RRCDA  
 rectangle *ABCD is a rectangle*  
 The definition of REABCD  
 is: RRDAB RRABC RRBCD RRCDA CRACBD

## Axioms of betweenness and congruence

betweennessidentity  
  hypotheses: none  
  conclusion: NOBEABA

betweennesssymmetry  
  hypotheses: BEABC  
  conclusion: BECBA

innertransitivity  
  hypotheses: BEABD BEBCD  
  conclusion: BEABC

connectivity  
  hypotheses: BEABD BEACD NOBEABC NOBEACB  
  conclusion: EQBC

nullsegment1  
  hypotheses: EEABCC  
  conclusion: EQAB

nullsegment2  
  hypotheses: none  
  conclusion: EEAABB

5-line  
  hypotheses: EEBCbc EEADad EEbDbd BEABC BEabc EEABab  
  conclusion: EEDCdc

## Postulates

extension  
  hypotheses: NEAB NECD  
  conclusion: For some X, BEABX EEbXCD

Pasch-inner  
  hypotheses: BEAPC BEBQC NCACB  
  conclusion: For some X, BEAXQ BEbXP

Pasch-outer  
  hypotheses: BEAPC BEBQC NCBQA  
  conclusion: For some X, BEAXQ BEbPX

line-circle  
  hypotheses: CIKCPQ ICBK NEAB  
  conclusion: For some XY, COABX COABY ONXK ONYK  
  BEXBY

circle-circle  
  hypotheses: CIJCRS ICPJ OCQJ CIKDFG ONPK ONQK  
  conclusion: For some X, ONXJ ONXK

Euclid5  
  hypotheses: BErts BEptq BEraq EEptqt EEtrts NCpqs  
  conclusion: For some X, BEpaX BEsqX

## Axioms for Equal Figures

Efreflexive  
  hypotheses: none

```

conclusion: EFABCDABCD
congruentequal
  hypotheses: TCABCabc
  conclusion: ETABCabc
ETpermutation
  hypotheses: ETABCabc
  conclusion: ETABCbca ETABCacb ETABCbac ETABCcba
  ETABCcab
ETsymmetric
  hypotheses: ETABCabc
  conclusion: ETabcABC
EFpermutation
  hypotheses: EFABCDabcd
  conclusion: EFABCDbcda EFABCDdcba EFABCDcdab EFABCDbadc
  EFABCDdabc EFABCDcbad EFABCDadcb
halvesofequal
  hypotheses: ETABCBCD OSABCD ETabcbcd OSabcd EFABDCabdc
  conclusion: ETABCabc
EFsymmetric
  hypotheses: EFABCDabcd
  conclusion: EFabcdABCD
EFtransitive
  hypotheses: EFABCDabcd EFabcdPQRS
  conclusion: EFABCDPQRS
ETtransitive
  hypotheses: ETABCabc ETabcPQR
  conclusion: ETABCPQR
cutoff1
  hypotheses: BEABC BEabc BEEDC BEedc ETBCDbcd ETACEace
  conclusion: EFABDEabde
cutoff2
  hypotheses: BEBCD BEbcd ETCDEcde EFABCEabde
  conclusion: EFABCEabce
paste1
  hypotheses: BEABC BEabc BEEDC BEedc ETBCDbcd EFABDEabde
  conclusion: ETACEace
deZolt1
  hypotheses: BEBED
  conclusion: NOETDBCEBC
deZolt2
  hypotheses: TRABC BEBEA BEBFC
  conclusion: NOETABCEBF
paste2
  hypotheses: BEBCD BEbcd ETCDEcde EFABCEabce
  conclusion: EFABDEabde
paste3
  hypotheses: ETABCabc ETABDabd OSCABD OScabd
  conclusion: EFACBDacbd
paste4
  hypotheses: EFABmDFKHG EFDBeCGHML OSABDC BEKHM BEFGL BEBmD BEBeC

```

conclusion: EFABCDFKML

## References

- [1] J. AVIGAD, E. DEAN, AND J. MUMMA, *A formal system for Euclid's Elements*, Review of Symbolic Logic, 2 (2009), pp. 700–768.
- [2] M. BEESON, *A constructive version of Tarski's geometry*, Annals of Pure and Applied Logic, 166 (2015), pp. 1199–1273.
- [3] ———, *Constructive geometry and the parallel postulate*, Bulletin of Symbolic Logic, 22 (2016), pp. 1–104.
- [4] W. W. BLEDSOE AND D. LOVELAND, *Automated Theorem Proving: After 25 Years*, American Mathematical Society, Providence, R. I., 1983.
- [5] P. BOUTRY, *On the Formalization of Foundations of Geometry*, PhD thesis, University of Strasbourg, in preparation.
- [6] P. BOUTRY, G. BRAUN, AND J. NARBOUX, *From Tarski to Descartes: Formalization of the Arithmetization of Euclidean Geometry*, in SCSS 2016, the 7th International Symposium on Symbolic Computation in Software Science, J. H. Davenport and F. Ghourabi, eds., vol. 39 of EPiC Series in Computing, Tokyo, Japan, Mar. 2016, EasyChair, pp. 14–28.
- [7] ———, *Formalization of the Arithmetization of Euclidean Plane Geometry and Applications*, Journal of Symbolic Computation, (2017), p. 23. to appear.
- [8] P. BOUTRY, C. GRIES, J. NARBOUX, AND P. SCHRECK, *Parallel postulates and continuity axioms: a mechanized study in intuitionistic logic using Coq*, Journal of Automated Reasoning, (2017), p. 68.
- [9] G. BRAUN, P. BOUTRY, AND J. NARBOUX, *From Hilbert to Tarski*, in Eleventh International Workshop on Automated Deduction in Geometry, Proceedings of ADG 2016, Strasbourg, France, June 2016, p. 19.
- [10] G. BRAUN AND J. NARBOUX, *From Tarski to Hilbert*, in Automated Deduction in Geometry 2012, T. Ida and J. Fleuriot, eds., vol. 7993, Edinburgh, United Kingdom, Sept. 2012, Jacques Fleuriot, Springer, pp. 89–109.
- [11] ———, *A synthetic proof of Pappus' theorem in Tarski's geometry*, Journal of Automated Reasoning, 58 (2017), p. 23.
- [12] S. C. CHOU, *Mechanical Geometry Theorem Proving*, Kluwer Academic Publishers, Norwell, MA, 1987.
- [13] ———, *An introduction to Wu's method for mechanical theorem proving in geometry*, Journal of Automated Reasoning, 4 (1988), pp. 237–267.
- [14] S.-C. CHOU, X.-S. GAO, AND J.-Z. ZHANG, *Machine Proofs in Geometry: Automated Production of Readable Proofs for Geometry Theorems*, World Scientific, 1994.



- [15] S. C. CHOU AND W. F. SCHELTER, *Proving geometry theorems with rewrite rules*, Journal of Automated Reasoning, 2 (1986), pp. 253–273.
- [16] P. CORBINEAU, *A declarative language for the coq proof assistant*, in Types for Proofs and Programs, vol. 4941 of LNCS, Springer, 2008.
- [17] EUCLID, *The Elements of Euclid, viz. the first six books, together with the eleventh and the twelfth*, Nourse and Balfous, Edinburgh, 7th ed., 1787. Translated by Robert Simson. Available from Bibliotheque Nationale at <http://gallica.bnf.fr/ark:/12148/bpt6k1163221v>.
- [18] ———, *The Thirteen Books of The Elements*, Dover, New York, 1956. Three volumes. Includes commentary by the translator, Sir Thomas L. Heath.
- [19] H. GELERNTER, *Realization of a geometry theorem-proving machine*, in Computers and Thought, E. Feigenbaum and J. Feldman, eds., McGraw-Hill, New York, 1963, pp. 134–152.
- [20] H. GELERNTER, J. R. HANSEN, AND D. W. LOVELAND, *Empirical explorations of a geometry-theorem proving machine*, in Computers and Thought, E. Feigenbaum and J. Feldman, eds., McGraw-Hill, New York, 1963, pp. 153–167.
- [21] M. J. GREENBERG, *Euclidean and non-Euclidean Geometries: Development and History*, W. H. Freeman, New York, fourth ed., 2008.
- [22] C. GRIES, *Axiomes de continuité en géométrie neutre : une étude mécanisée en Coq*, September 2017. University of Strasbourg, internship report.
- [23] H. N. GUPTA, *Contributions to the Axiomatic Foundations of Geometry*, PhD thesis, University of California, Berkeley, 1965.
- [24] R. HARTSHORNE, *Geometry: Euclid and Beyond*, Springer, New York, 2000.
- [25] D. HILBERT, *Foundations of Geometry (Grundlagen der Geometrie)*, Open Court, La Salle, Illinois, 1960. Second English edition, translated from the tenth German edition by Leo Unger. Original publication date, 1899.
- [26] D. KAPUR, *A refutational approach to geometry theorem proving*, Journal of Artificial Intelligence, 37 (1988), pp. 61–93.
- [27] B. KUTZLER AND S. STIFTER, *Automated geometry theorem proving using Buchberger’s algorithm*, in Proceedings of the International Symposium on symbolic and algebraic computation (SYMSAC 86), Waterloo, 1986, B. Char, ed., 1986, pp. 209–214.
- [28] J. MOLLERUP, *Die Beweise der ebenen Geometrie ohne Benutzung der Gleichheit und Ungleichheit der Winkel*, Mathematische Annalen, 58 (1904), pp. 479–496.
- [29] J. NARBOUX, P. JANIČIĆ, AND J. FLEURIOT, *Computer-assisted Theorem Proving in Synthetic Geometry*. unpublished, 2017.
- [30] M. PASCH, *Vorlesung über Neuere Geometrie*, Teubner, Leipzig, 1882.

- [31] G. PEANO, *Principii de Geometria*, Fratelli Bocca, Torino, 1889.
- [32] R. POTTS, *Euclid's Elements of geometry [Books 1-6, 11,12] with explanatory notes; together with a selection of geometrical exercises. To which is prefixed an intr., containing a brief outline of the history of geometry.*, Oxford University Press, 1845.
- [33] PROCLUS, *A commentary on the first book of Euclid*, Princeton University Press, 1970.
- [34] A. QUAIFFE, *Automated development of Tarski's geometry*, Journal of Automated Reasoning, 5 (1989), pp. 97–118.
- [35] W. SCHWABHÄUSER, W. SZMIELEW, AND A. TARSKI, *Metamathematische Methoden in der Geometrie: Teil I: Ein axiomatischer Aufbau der euklidischen Geometrie. Teil II: Metamathematische Betrachtungen (Hochschultext)*, Springer-Verlag, 1983. Reprinted 2011 by Ishi Press, with a new foreword by Michael Beeson.
- [36] S. STOJANOVIC DURDEVIC, J. NARBOUX, AND P. JANICIC, *Automated Generation of Machine Verifiable and Readable Proofs: A Case Study of Tarski's Geometry*, Annals of Mathematics and Artificial Intelligence, 73 (2015), p. 25.
- [37] J. STROMMER, *Über die Kreisaxiome*, Periodica Mathematica Hungarica, 4 (1973), pp. 3–16.
- [38] A. TARSKI AND S. GIVANT, *Tarski's system of geometry*, The Bulletin of Symbolic Logic, 5 (1999), pp. 175–214.
- [39] G. VERONESE, *Fondamenti di geometria a più dimensioni e a più specie di unità rettilinee esposti in forma elementare. Lezioni per la scuola di magistero in matematica*, Padova: Tipografia del Seminario, 1891, 1891.
- [40] W.-T. WU, *On the decision problem and the mechanization of theorem-proving in elementary geometry*, Scientia Sinica, 21 (1978), p. reprinted in [4].
- [41] ———, *Mechanical Theorem Proving in Geometries: Basic Principles*, Springer-Verlag, Wien/ New York, 1994.