Proof-checking Euclid

Euclides ab omni naevo vindicatus

Michael Beeson
joint work with
Freek Wiedijk
Julien Narboux

profbeeson@gmail.com

February 27, 2018

Proof-checking Euclid

Euclides ab omni naevo vindicatus

Michael Beeson
joint work with
Freek Wiedijk
Julien Narboux

profbeeson@gmail.com

February 27, 2018

from Rafael's *School of Athens*

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*.
- Abraham Lincoln claimed to have read the first six books of Euclid thoroughly and learned from them how to "demonstrate" something in court. He referred to Euclid repeatedly in the Lincoln-Douglas debates.
- The *Elements* also inspired the form of the American Declaration of Independence. Jefferson continued reading Euclid all his life.

Flaws in Euclid, and fixing them

- Proclus (AD 150) already had complaints.
- Starting in the 1700s, 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.
- That led to more care with proofs and eventually to non-Euclidean geometry.
- In the modern era, beginning already in the nineteenth century, the standards of proof in mathematics became more demanding. Italian and German geometers studied the axioms carefully.
- Pasch, Pieri, and Peano made fundamental contributions, especially the need for the notion "between" and axioms about it.

What is proof-checking?

- It is not (mere) calculation ("computer algebra").
- A (formal) proof is a sequence of symbols combined according to certain rules.
- Hilbert's "tables, chairs, and beer mugs" remark is relevant.
- The rules are simple and a computer can check if a given sequence of symbols does or does not follow the rules.
- Typically there are four to ten more steps in a formal proof than in a detailed textbook proof, as the inferences are smaller.
- Journal proofs of course are even less detailed.
- To construct a formal proof one usually proceeds interactively, as it's difficult not to make errors. Hence "ITP".
- "ATP" means "automatic" theorem proving. Get the computer to find the proof. ITP tries to use ATP to fill in the tiny steps.

State of the art in proof-checking

- The programs used for ITP are called "proof assistants".
- There is a book *The 17 Provers of the World* by my co-author Freek Wiedijk. Some of the 17 are ATP only; many are ITP.
- The most famous three are HOL Light, Coq, and Isabelle.
- Most of undergraduate mathematics has been formally checked. Each prover has "libraries" of already-proved mathematics.
- There is a list of 100 famous "big" theorems. People have been "knocking them off" for 20 years now. Most are proved.
- That list includes the four-color problem. Proof-checking the four-color problem is quite different than using a computer to make calculations in hundreds of different cases.
- In 2013 an international collaboration of hundreds of mathematicians and computer scientists completed proof-checking the Kepler Conjecture, which is now without a doubt Hales's theorem.

Proof-checking Euclid

- Up to now, nobody has tried to check the proofs of Euclid.
- It is our opportunity. We happen to exist at the historical moment when proof-checking has come of age.

Aims of this work

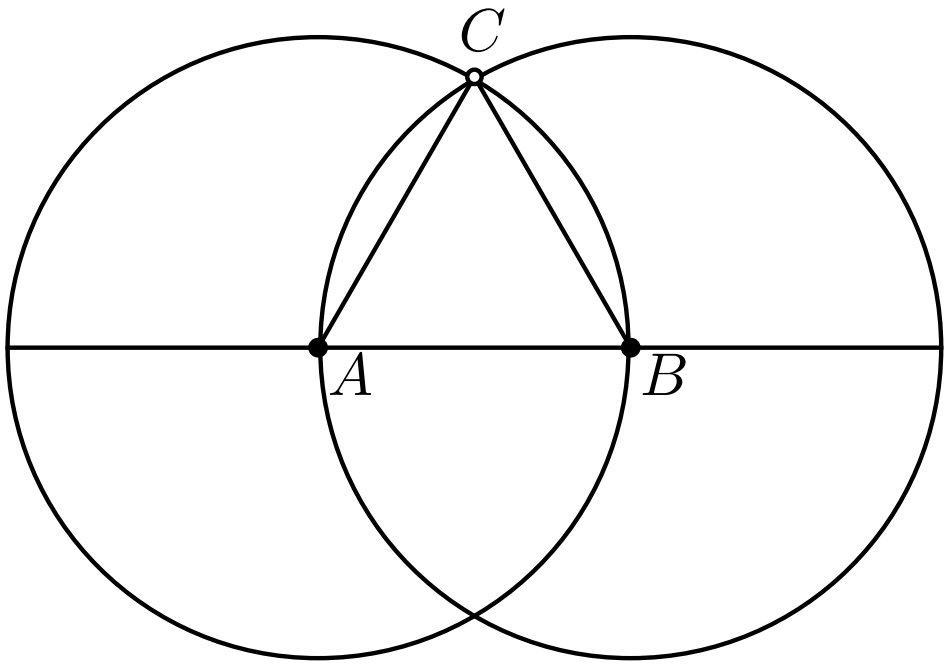
- Fix Euclid's axioms (and common notions), using an axiom system rather close to Euclid's, but including some axioms about betweenness and circles that Euclid neglected to state.
- Give correct proofs of all the propositions in Book I from the new axioms, following Euclid's proofs as closely as possible.
- Show that those proofs are indeed correct by checking the proofs using the proof-checking programs HOL Light and Coq.

Proof-checking Euclid

There is some bad news and some good news:

- Euclid's axioms are imprecise and insufficient.
- Euclid's proofs have gaps and outright errors.
- Euclid's definitions are fairly precise.
- Euclid's proofs are fairly detailed.

Proposition I.1, equilateral triangle construction

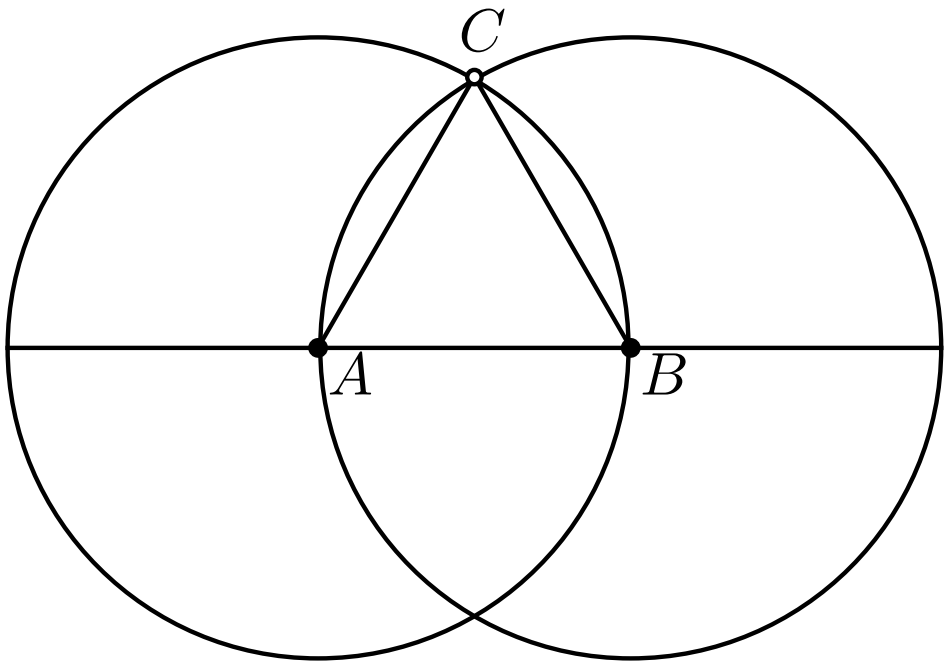


Why do the circles meet at C?

None of Euclid's postulates guarantees it.

The problem here is a missing axiom, not a mistake in reasoning.

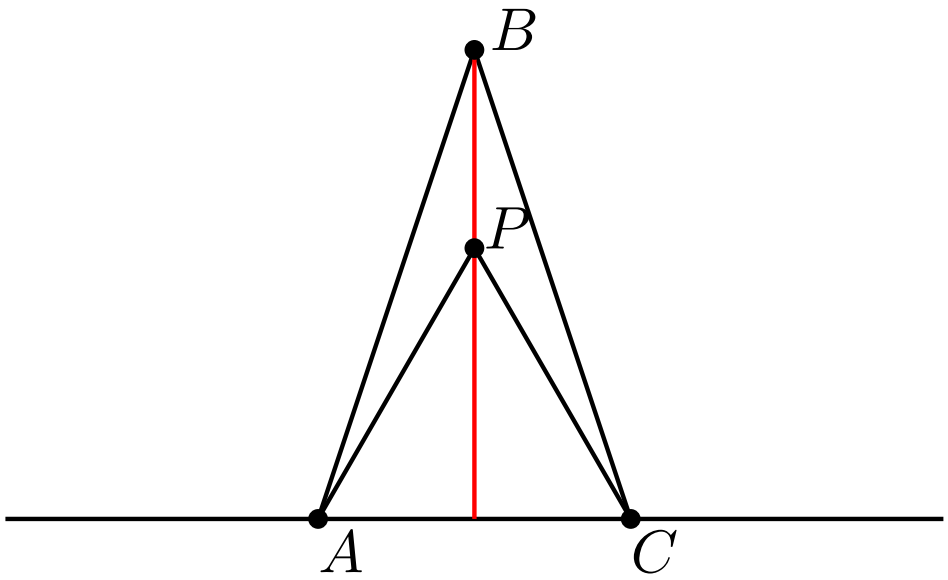
Proposition I.1, a second blemish



Why do the three points form a triangle? That is, why are they not collinear?

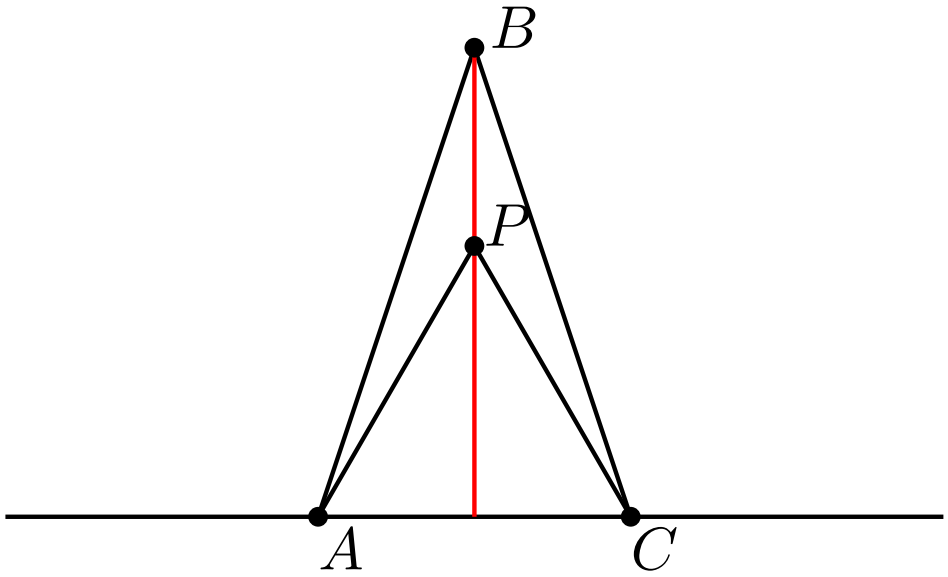
- If not then one of them must be between the other two.
- We need: if B is between A and C then AB and AC are not equal.
- So this theorem should come before Euclid's first proposition.

Prop. I.9, Euclid's bisection of angle ABC



- Construct P by I.1 so ACP is equilateral.
- Then BP is the bisector.
- Oops! What if B and P are the same point?
- Then could we just take "the other equilateral triangle" ?

Prop. I.9 continued



- Prop I.1 only claimed *one* equilateral triangle, not two.
- To get two we would need a stronger circle-circle axiom.
- Anyway even if we did that there is another difficulty: Why is the red line in the interior of the angle? That is, why does the red line meet AC?
- Euclid wishes to use I.9 to *conclude* in I.10 that the red line meets AC. But he has not proved it in I.9.

There is no dimension axiom in Euclid.

- Euclid Book XI is about three-dimensional geometry. He wants to develop the theory of the Platonic solids.
- Therefore Euclid is *not* meant to be about plane geometry only.
- For example the definition of parallel lines is, lines that lie in the same plane and do not meet.
- Therefore every theorem should be true in 2-space and in 3-space.
- In 3-space, "circles" are really "spheres".
- Therefore on AB there might be not just two but infinitely many equilateral triangles.

Fix I.9 by using a strong circle-circle axiom?

- To do so, we would have to assume that two circles meet in two points P and Q such that the line PQ meets the line joining the centers of the circle. That amounts to assuming that segments and angles can both be bisected.
- If we had no other way to prove I.9 and I.10 but to assume them, then we would have had to do it that way.
- If this problem was pointed out in the literature before 2016, I do not know where.
- Luckily, we can prove I.9 and I.10 another way

Gupta's midpoint theorems

- Tarski's student Gupta earned his Ph. D. at Berkeley in 1965. His amazing thesis contained enough material for at least three theses.
- His most amazing result is that it is possible to construct both dropped and erected perpendiculars *without using circles at all*.
- This result goes far beyond Euclid in difficulty and we do not need it.
- Instead we use his simple and beautiful proof that the base of an isosceles triangle has a midpoint.
- That gives us I.10, the bisection of a segment, and from that we get I.9, the bisection of an angle.
- So I.10 and I.9 have to be proved in a different order than Euclid claimed.
- Before 1965, the "proofs" of I.9 and I.10 were irremediably defective. This is more than a "blemish."

Euclid's Axioms

We now discuss the axioms of geometry. First we review Euclid's framework.

- Primitive notions. Euclid tries to define *everything*, whereas today, we wish to start with a few undefined notions.
- Euclid never *used* his definitions of point and line, in effect taking those notions as undefined.
- Common notions. These amount, in today's terms, to equality axioms, plus things like "the whole is equal to the sum of the parts" and "the whole is not equal to the part." The word "equal" is used for points, lines, angles, and "figures" (polygons).
- Definitions. These look fairly modern.
- Postulates. Euclid had five.

Euclid's postulates

- To draw a straight line from any point to any point.
- To produce a finite straight line continuously in a straight line.
- To describe a circle with any center and radius.
- Euclid 4: All right angles are equal
- Euclid 5: if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

What Euclid omitted

- Betweenness, written $B(a, b, c)$, meaning that a, b, and c are distinct points on a line with b between a and c.
- Pasch's axiom (or axioms) about betweenness, introduced by Pasch in 1882 and refined by Peano in 1889.
- Other axioms about betweenness, about the order of points on a line.
- An axiom permitting a proof of the SAS congruence criterion, which Euclid tried unsuccessfully to prove in I.2.
- A line-circle axiom is also needed (in addition to circle-circle).
- Since equality of lines is not really equality, the common notions don't really cover the required axioms about equality of lines. In other words, those notions are not "common" enough to remain unstated.

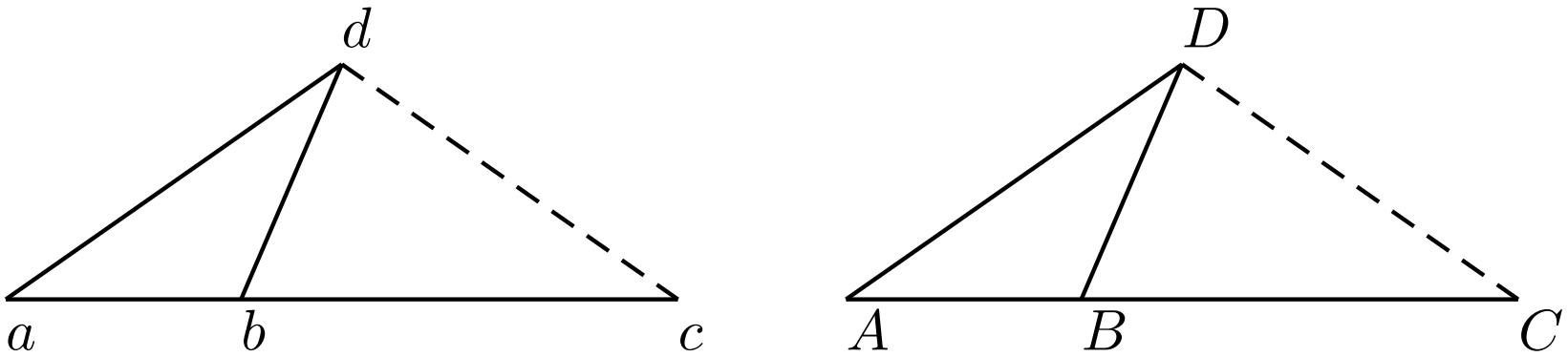
Our framework in which to formalize Euclid

- A first-order theory with variables of two sorts: points and circles.
- Angles are treated as given by three points. Thus, like Euclid, we speak of "angle ABC".
- Lines are treated as given by two points, as in line AB. As in Euclid, lines always have endpoints, but can be extended.
- The primitive relations are "congruence" and betweenness.
- Congruence is a 4-ary relation between points, written $AB = CD$. We follow Euclid in using the word "equality" instead of "congruence".
- We want to have variables for circles, even though Euclid did not. Euclid repeatedly fell into error by naming a circle by three points, such as ABC, where point C has not really been proved to exist. are repaired by saying "let K be the circle with center C and radius AB."

Our axioms (selected by us, but at least a century old)

- Betweenness axioms about the order of points on a line. We call these the "linear betweenness axioms."
- Congruence axioms, including a line-extension axiom.
- Two forms of the Pasch axiom, "inner Pasch" and "outer Pasch", introduced by Peano in 1890.
- Tarski's 5-segment axiom, which enables us to prove SAS.
- line-circle and circle-circle
- Euclid 5
- We do *not* have Euclid 4 (all right angles equal) as it can be proved.
- Thus we have supplied what Euclid omitted, *and nothing more*.

The 5-line axiom



If the four solid lines on the left are equal to the corresponding solid lines on the right, then the dashed lines are also equal.

- The conclusion is, in effect, the congruence of triangles dbc and DBC.
- The 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.
- The 5-line axiom is a points-only version of the SAS triangle congruence theorem.

History of the five-line axiom

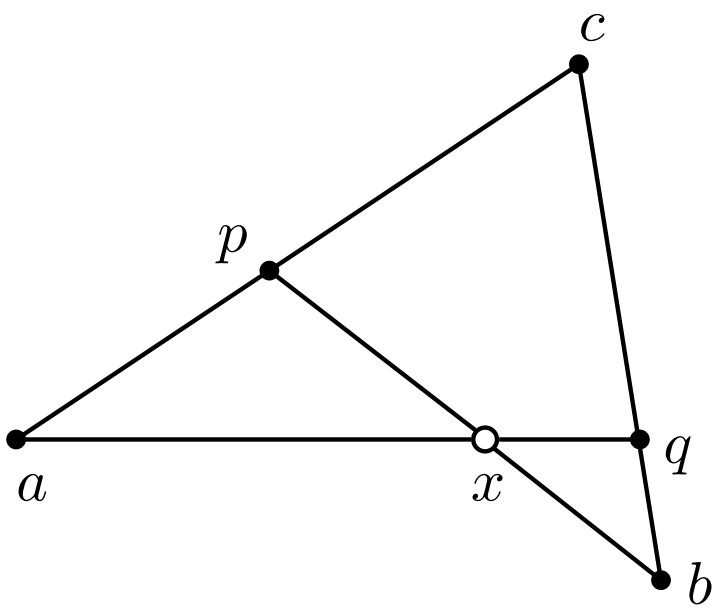
- Our version of the five-line axiom was introduced by Tarski (in 1927), although we have changed non-strict betweenness to strict betweenness.
- The key idea (replacing reasoning about angles by reasoning about congruence of segments) was introduced (in 1904) by J. Mollerup.
- Mollerup (without comment) gives a reference to Veronese 1891.
- 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.

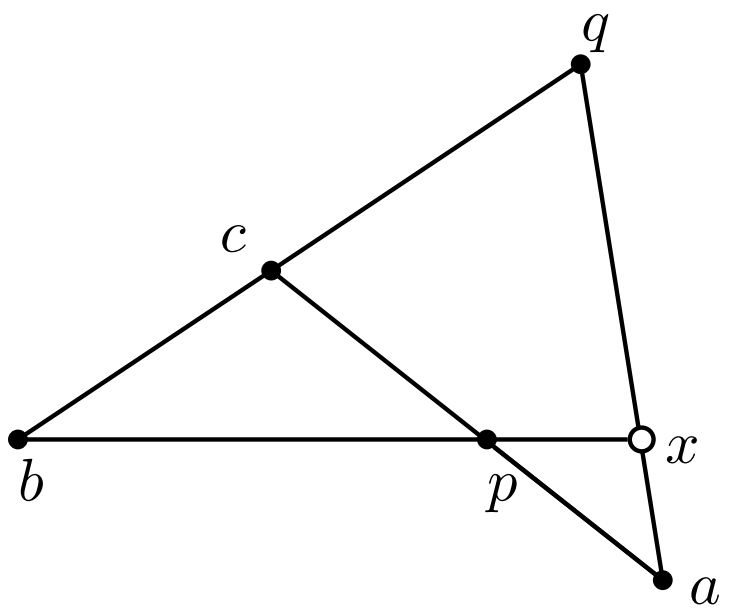
Pasch's axiom and Peano's improved versions

- Pasch 1882 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 is only true in a plane, which is not to be assumed in Euclid.
- Seven years later, Peano introduced what are now called "inner Pasch" and "outer Pasch", which work without a dimension axiom.
- Peano wrote everything in formal symbols only, and eventually bought his own printing press to print his books himself.

Inner and outer Pasch

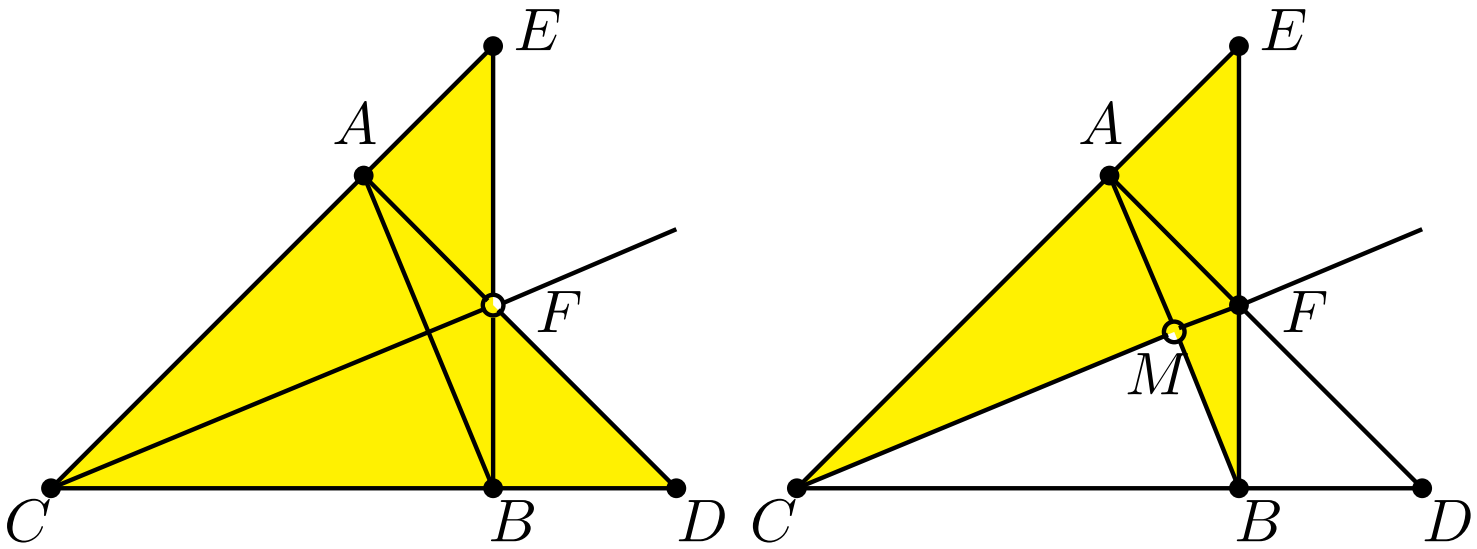
Line pb meets triangle acq in one side ac, and meets an extension of side cq. Then it also meets the third side aq. The open circles show the points asserted to exist.





Gupta's "little midpoint theorem"

We show how Gupta used inner Pasch to construct the midpoint M of line AB, given AB is the base of an isosceles triangle ABC. Namely, construct first F and then M by inner Pasch.

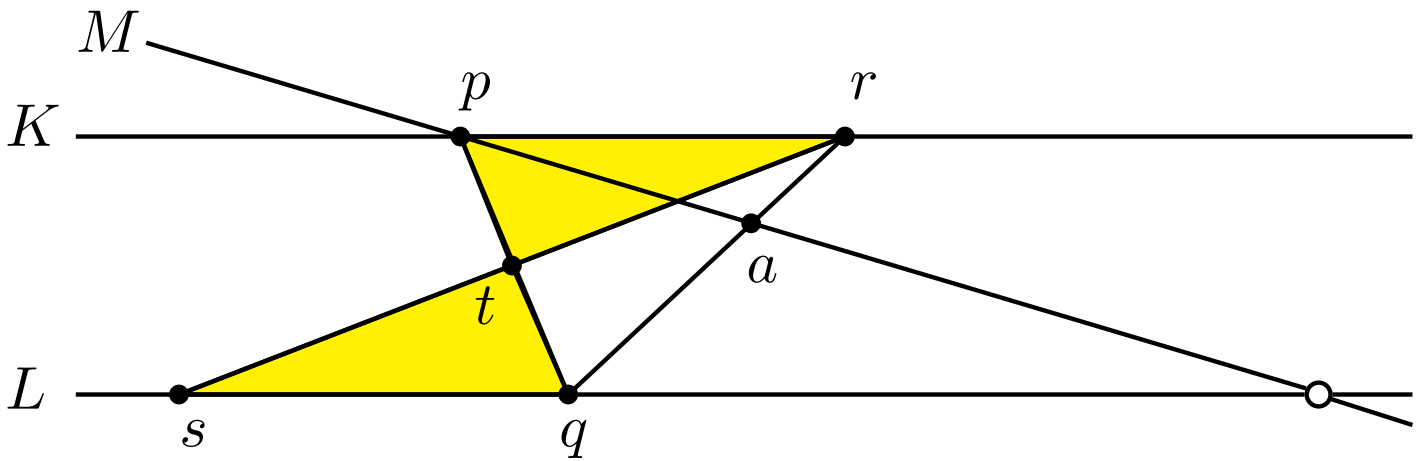


Then $AD = BE$ by the 5-line axiom applied to $ACBD$ and $BCAE$. (The proof is by no means done; we do not finish it here.)

Euclid's Postulate 5, the "parallel postulate"

We give a "points only" version, not mentioning angles.

Transversal pq of lines M and L makes corresponding interior angles less than two right angles, as witnessed by a. That is expressed simply by $\mathbf{B}(r, a, q)$. The shaded triangles are assumed congruent. That is, their corresponding sides are equal. Then M meets L as indicated by the open circle.



Equal Figures in Euclid

- Euclid used the word "figure" to mean what we now call a simple closed polygon.
- 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."

Formalizing "equal figures"

- Some textbooks try to define "equal figures" as "figures with equal area."
- Euclid probably did not do that because he did not know how to define "area".
- We do not do that since defining area requires segment arithmetic, which takes us too far from Euclid.
- In effect Euclid axiomatizes the notion. So do we.
- Euclid appealed to the common notions, saying "the whole is equal to the sum of the parts" when he probably meant, "the area of the whole is the sum of the areas of the parts."
- Similarly, "the part is not equal to the whole" for "the area of the part is not equal to the area of the whole."
- But Euclid couldn't define "area" so he skirted the issue.
- The theory of segment arithmetic has been formalized in Coq already, so we could use that, but instead we follow Euclid, stating explicitly the axioms he uses.

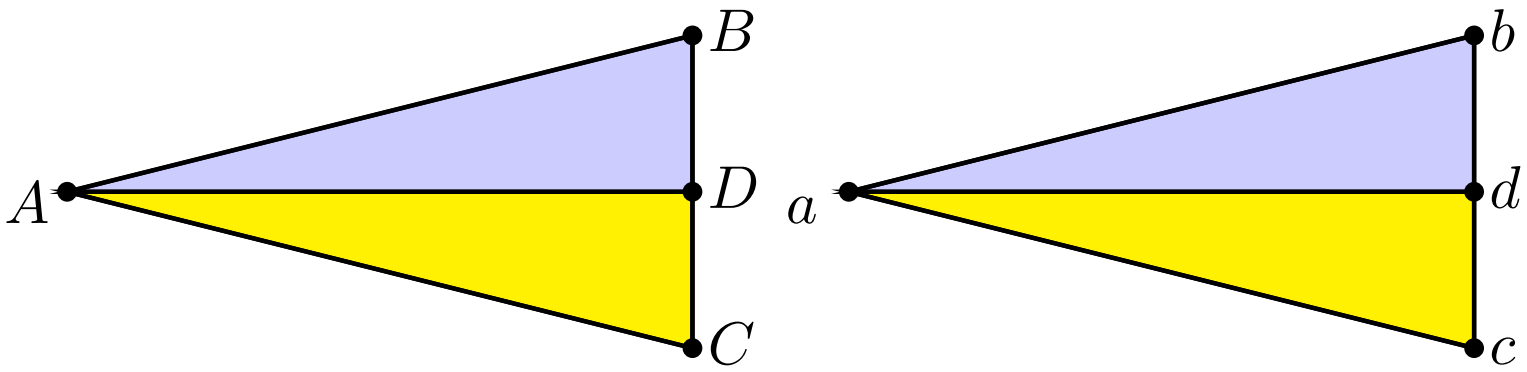
Formalizing "equal figures"

- Axiomatizations in the literature either use real numbers or natural numbers. We gave a first-order axiomatization.
- We used `ETABCabc` to express "ABC and abc are equal triangles.
- We used `EFABCDabcd` to express "$ABCD$ and $abcd$ are equal quadrilaterals.

Two sample equal-figure axioms

If ABC is equal to abc then BCA is equal to bca. This is `ETpermutation`. It is closely related to the principle that the area of a triangle, defined as half the base times the altitude, doesn't depend on which side is considered the base.

If CAD is equal to cad, and DAB is equal to dab, then CAB is equal to cab. This is `ETpaste1`.



There are altogether 18 equal-figure axioms.

"Lie in the same plane"

- Euclid defined *parallel lines* to be lines that lie in the same plane and do not meet.
- But he failed to define "lie in the same plane".
- Tarski defined "A and B lie on the same side of line L" as shown below.
- This definition works without a dimension axiom and shows that a point and a line determine two half-planes.

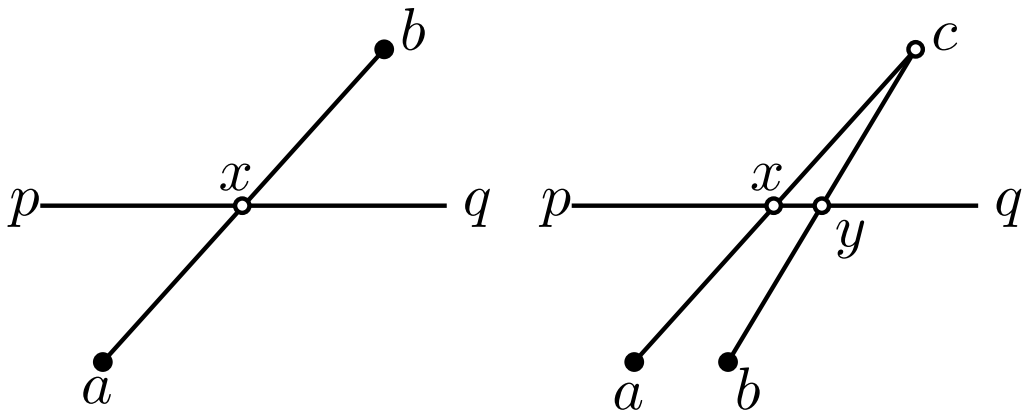


Figure : (Left) a and b are on the opposite side of pq. (Right) a and b are on the same side of pq if there exist points x and y collinear with pq, and a point c, such that $\mathbf{B}(a, x, c)$ and $\mathbf{B}(b, y, c)$.

Formalizing Euclid's proofs

Different ways of writing the same formula are used in English, TEX, HOL Light, Coq, and our proof debugger. Simple scripts can translate one to the other.

- B is between A and C
- $\mathbf{B}(A, B, C)$
- $BE(A, B, C)$
- `BEABC`
- A, B, and C are collinear
- `COABC`
- several dozen two-letter function symbols are defined, following Euclid.

Start of the proof of I.10

```
NEAB
ANELABC+TRABC  proposition:01
ELABC
TRABC
NCABC   defn:triangle
ANEEABBC+EEBCCA  defn:equilateral
EEBCCA
EEACCB  lemma:doublereverse
EEACBC  lemma:congruenceflip
EQCB  assumption
 COACB defn:collinear
 COABC lemma:collinearorder
NECB   reductio
ANBECBD+EEBDAB  postulate:extension
EEBDAB
```

Structure of Euclid's proofs

- Each line contains a statement, either a literal or a conjunction or disjunction of literals.
- Proof by reductio and proof by cases are allowed.
- Each line is either justified or is an assumption (for reductio or cases), or is a repeat of a previous line.
- initial unjustified lines repeat the hypotheses of the theorem.

Writing and debugging the proofs

- We kept a "master list" of all the theorems in the order that they need to be proved.
- Also we kept lists of all the axioms, postulates, and definitions.
- We wrote a custom proof-checker or proof debugger to check all the proofs in the proper order.
- We used Polish notation so formulas can be represented as strings. Euclid has no function symbols.
- Therefore we did not have to write a parser.
- Unification is easy to write using regular expressions for matching, when formulas are strings.
- Since we wrote this checker, we could customize the output when checking failed to provide maximum help in understanding what failed.
- This checker may have bugs. If it does, that does not affect the validity of our results, because we used it only as a debugger, to prepare proofs for more reliable checkers.

Two hundred proofs

- Our formalization consists of more than two hundred such proofs, including proofs of the 48 propositions of Book I.
- The proof-debugger runs in a few seconds. Here is the last part of the output:

```
Checking rectanglereverse.prf
Proof checked OK. 10 inferences. That makes 227 proofs.
Checking rectanglerotate.prf
Proof checked OK. 5 inferences. That makes 228 proofs.
Checking squaresequal.prf
Proof checked OK. 53 inferences. That makes 229 proofs.
Checking paste5.prf
Proof checked OK. 56 inferences. That makes 230 proofs.
Checking Prop48A.prf
Proof checked OK. 91 inferences. That makes 231 proofs.
Checking Prop48.prf
Proof checked OK. 59 inferences. That makes 232 proofs.
```

What are the other 174 proofs?

- Book Zero: Preliminaries of a very fundamental nature. For example, the plane separation theorem.
- Propositions that Euclid omitted but were used implicitly.
- The theory of angles; equality and order of angles are defined, not taken as undefined, and their properties have to be proved. For example, it is hard to prove that an angle cannot be less than itself.
- Advanced theorems that we found necessary to fill Euclid's gaps. For example, every square is a parallelogram.
- Some convenient variants of Euclid's propositions.
- The proof of Euclid 4 (all right angles are equal).

Conclusions

- We fixed Euclid's axioms, by adding inner and outer Pasch, the 5-line axiom, the linear betweenness axioms, line-circle and circle-circle.
- We fixed Euclid's treatment of equal figures, by making the equal-figure axioms explicit.
- After filling the gaps caused by missing axioms, there are still serious errors in Euclid.
- We were able to fix those by supplying some fundamental theorems in "Book Zero", some missing arguments in the proofs of his propositions, and some theorems Euclid used implicitly, and should have proved (such as, every square is a parallelogram).
- The result is a clean and completely rigorous treatment of Euclid Book I, following the lines of Euclid's original as closely as is possible, while still fixing the problems of the original.