

The Role of the Diagram in Euclid's *Elements*

Jeremy Avigad

Department of Philosophy and Department of Mathematical Sciences
Carnegie Mellon University

(including joint work with Ed Dean, John Mumma, and Ben Northrop)

June, 2008

The *Elements*

For more than two thousand years, Euclid's *Elements* was held to be the paradigm for rigorous argumentation.

But the nineteenth century raised concerns:

- Conclusions are drawn from diagrams, using “intuition” rather than precise rules.
- Particular diagrams are used to infer general results (without suitable justification).

Axiomatizations due to Pasch and Hilbert, and Tarski's formal axiomatization later on, were thought to make Euclid rigorous.

The *Elements*

But in some ways, they are unsatisfactory.

- Proofs in the new systems look very different from Euclid's.
- The initial criticisms belie the fact that Euclidean practice was remarkably stable for more than two thousand years.

Our project (Mumma, Dean, and me):

- Describe a formal system that is much more faithful to Euclid.
- Argue that the system is sound and complete (for the theorems it can express) relative to Euclidean fields.
- Show that the system can easily be implemented using contemporary automated reasoning technology.

Proposition 10

To bisect a given finite straight line.

Let AB be the given finite straight line.

Thus it is required to bisect the finite straight line AB .

Let the equilateral triangle ABC be constructed on it, [I. 1]
and let the angle ACB be bisected by the straight line CD ;

[I. 9]

I say that the straight line AB has been bisected at the point D .

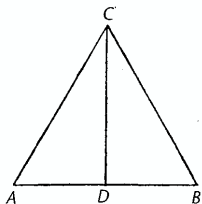
For, since AC is equal to CB , and CD is common,

the two sides AC , CD are equal to the two sides BC , CD respectively;
and the angle ACD is equal to the angle BCD ;

therefore the base AD is equal to the base BD .

[I. 4]

Therefore the given finite straight line AB has been bisected at D .



Q.E.F.

Proposition 16

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

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

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

Let AC be bisected at E , [I. 10]

and let BE be joined and produced in a straight line to F ;

let EF be made equal to BE , [I. 3]

let FC be joined, [Post. 1]

and let AC be drawn through to G . [Post. 2]

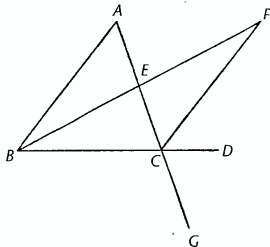
Then, since AE is equal to EC , and BE to EF ,
the two sides AE , EB are equal to the two sides
 CE , EF respectively;

and the angle AEB is equal to the angle FEC , for they are vertical angles. [I. 15]

Therefore the base AB is equal to the base FC , and the triangle ABE is equal to the triangle CFE ,

and the remaining angles are equal to the remaining angles respectively, namely those which the equal sides subtend; [I. 4]

therefore the angle BAE is equal to the angle ECF .



First salient feature: the use of diagrams

Observation: the diagram is inessential to the communication of the proof. (Rather, it is used to “see” that the inferences are correct.)

Exercise:

- Let p and q be points on a line.
- Let r be between p and q .
- Let s be between p and r .
- Let t be between r and q .

Is s necessarily between p and t ?

Philosophical thesis: from a logical perspective, the right way to characterize diagrammatic reasoning is in terms of the class of inferences that are licensed.

First salient feature: the use of diagrams

Observation: the diagram is inessential to the communication of the proof. (Rather, it is used to “see” that the inferences are correct.)

Exercise:

- Let p and q be points on a line.
- Let r be between p and q .
- Let s be between p and r .
- Let t be between r and q .



Is s necessarily between p and t ?

Philosophical thesis: from a logical perspective, the right way to characterize diagrammatic reasoning is in terms of the class of inferences that are licensed.

First salient feature: the use of diagrams

Observation (Manders): In a Euclidean proof, diagrams are only used to infer “co-exact” (regional / topological) information, such as incidence, intersection, containment, etc.

Exact (metric) information, like congruence, is always made explicit in the text.

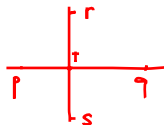
Poincaré: “Geometry is the art of precise reasoning from badly constructed diagrams.”

Solution: take the “diagram” to be a representation of the relevant data.

Second salient feature: generality

Some aspects of diagrammatic inference are puzzling:

- Let p and q be distinct points.
- Let L be a line through p and q .
- Let r and s be points on opposite sides of L .
- Let M be the line through r and s .
- Let t be the intersection of L and M .



Is t necessarily between r and s ? Is t necessarily between p and q ?

A diagrammatic inference was necessary even to “see” the L and M necessarily intersect. But not every feature found in a particular diagram is generally valid.

Euclid manages to avoid drawing invalid conclusions.

We need an explanation as to what secures the generality.

Third salient feature: logical form

Theorems in Euclid are of the form:

Given points, lines, circles, satisfying . . . , there are points, lines, circles satisfying . . .

where each . . . is a conjunction of literals.

(If the inner existential quantifier is absent, it is a “demonstration” rather than a “construction.”)

Proofs contain a construction part, and a deduction part.

Reasoning is linear, assertions are literals.

Exceptions: proof by contradiction, using a case distinction (sometimes “without loss of generality”).

Fourth salient feature: nondegeneracy

In the statement of a theorem, points are generally assumed to be distinct, triangles are nondegenerate, etc.

Two issues:

- Sometimes the theorem still holds in some degenerate cases.
- When the theorems are applied, Euclid doesn't always check nondegeneracy.

I will have little to say about this; in our system, nondegeneracy requirement are stated explicitly.

Formalizing Euclid

Prior efforts:

- Nathaniel Miller's Ph.D. thesis (2001): system is very complicated; generality is attained by considering cases exhaustively.
- John Mumma's Ph.D. thesis (2006): employs diagrams (and equivalence relation on diagrams); generality is attained using rules.

Our formal system, E , is derived from Mumma's. But now a “diagram” is nothing more than an abstract representation of topological information. The system spells out what can be inferred from the diagram.

The language of E

Basic sorts:

- diagram sorts: points p, q, r, \dots , lines L, M, N, \dots , circles $\alpha, \beta, \gamma, \dots$
- metric sorts: lengths, angles, and areas.

Basic symbols:

- diagram relations: $\text{on}(p, L)$, $\text{same-side}(p, q, L)$, $\text{between}(p, q, r)$, $\text{on}(p, \gamma)$, $\text{inside}(p, \gamma)$, $\text{center}(p, \gamma)$, $\text{intersects}(L, M)$, $=$
- metric functions and relations: $+$, $<$, $=$, right – angle
- connecting functions: \overline{pq} , $\angle pqr$, Δpqr

Other relations can be defined from these; e.g.

$$\text{diff-side}(p, q, L) \equiv \neg\text{on}(p, L) \wedge \neg\text{on}(q, L) \wedge \neg\text{same-side}(p, q, L)$$

Sequents

The proof system establishes sequents of the following form:

$$\Gamma \Rightarrow \exists \vec{q}, \vec{M}, \vec{\beta}. \Delta$$

where Γ and Δ are sets of literals.

Applying a construction rule or prior theorem augments $\vec{q}, \vec{M}, \vec{\beta}, \Delta$.

Applying deductive inferences augments Δ .

Case splits and suppositional reasoning temporarily augment Γ .

I need to describe:

- Construction rules.
- Deductive inferences.

Diagram inferences are implicit in both.

Construction rules

“Let p be a point on L ”

No prerequisites.

“Let p be a point distinct from q and r ”

No prerequisites.

“Let L be the line through p and q ”

Requires $p \neq q$.

“Let p be the intersection of L and M .”

Requires that L and M intersect.

And so on. . .

Deductive inferences

Four types:

1. Diagram inferences: any fact that can be “read off” from the diagram.
2. Metric inferences: essentially linear arithmetic on lengths, angles, and areas.
3. Diagram to metric: for example, if q is between p and r , then $\overline{pq} + \overline{qr} = \overline{pr}$, and similarly for areas and angles.
4. Metric to diagram: for example, if p is the center of γ , q is on γ , and $\overline{pr} < \overline{pq}$, then r is inside γ .

Diagram inferences

Both construction inferences and diagram inferences require an account of what can be “read off” from the diagram.

We get this by closing the diagrammatic data in $\Gamma \cup \Delta$ under various rules, including:

- properties of “between”
- properties of “same side”
- “Pasch rules,” relating “between” and “same side”
- triple incidence rules
- circle rules
- intersection rules

These yield conclusions that are generally valid, that is, common to all possible realizations.

Completeness

Tarski's first-order axiomatization of Euclidean geometry yields a complete theory of the Euclidean plane (inter-interpretable with real closed fields).

Drop the completeness axiom, and replace it with an axiom asserting that if a line L passes through a point inside a circle α , then L and α intersect.

The resulting theory is inter-interpretable with the theory of "Euclidean fields," and so is complete wrt "ruler and compass constructions." (Ziegler: it is also undecidable.)

Theorem. If a sequent of E is valid wrt to ruler and compass constructions, it can be derived in E .

Completeness

One strategy: interpret Tarski's theory in E .

Problem: Tarski includes full first-order logic!

Solution: With slight tinkering, Tarski's theory can be made "geometric," i.e. the axioms can be put in a restricted logical form.

A cut-elimination theorem due to Negri and von Plato then implies that any geometric assertion provable in Tarski's theory has a geometric proof.

Such a proof can be simulated in E .

Completeness

Ed Dean is carrying out the details in this MS thesis:

1. Suppose a sequent A of E is valid for the intended semantics.
2. Then a translation $\sigma(A)$ to Tarki's language is valid for ruler and compass constructions.
3. So it is provable in Tarski's theory.
4. So it has a cut-free proof.
5. This proof can be translated back to E , so E proves $\tau(\sigma(A))$.
6. From this, E can derive the original sequent, A .

Implementation

Ben Northrop (working part-time towards his MS thesis) implemented the diagram inferences in Java, with a saturation algorithm.

For moderately complex diagrams, the implementation is too slow to be of practical use.

We also tried first-order theorem provers, Spass and E, which do well on the diagrammatic inferences (E does better).

Alas, our diagram axioms are universal, which puts them outside the SMT framework.

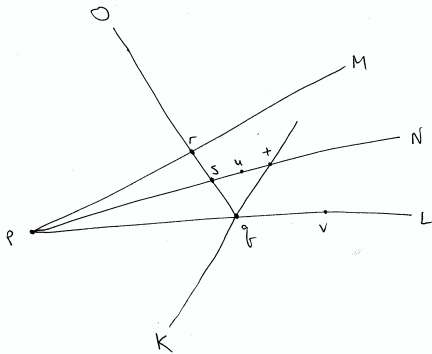
But some SMT provers use heuristic instantiation of quantifiers. . .

Implementation

In a fairly complicated diagram, Z3 plows through all our inferences automatically. CVC3 also does pretty well.

In fact, they also handle all the metric inferences, as well as the ordinary propositional logic needed to handle case splits, proofs by contradiction, and so on.

In other words, SMT provers can be used as a complete back end.



```

:formula (sameside p t O)
:formula (sameside s t O)
:formula (not (sameside s t M))
:formula (not (sameside u t M))
:formula (bet s p t)
:formula (= M N)
:formula (bet q s u)
:formula (on q N)
:formula (= q t)
:formula (not (< (seg s u) (seg s t)))
:formula (not (< (seg u s) (seg s t)))
:formula (not (< (+ (seg s u) (seg u t)) (seg p t)))
:formula (not (< (+ (seg u s) (seg u t)) (seg p t)))
:formula (on u L)
:formula (on t L)
:formula (on p K)
:formula (not (sameside r s L))
:formula (not (sameside s u L))
:formula (not (sameside r u L))
:formula (sameside s v K)
:formula (not (= (+ (angle r p s) (angle s p q)) (angle r p q)))
:formula (not (sameside p s K))
:formula (not (sameside s t L))
:formula (= L K)
:formula (= q s)
:formula (= q t)
:formula (= q p)
:formula (not (= (+ (angle p q s) (angle s q t)) (angle p q t)))
:formula (not (< (angle p q s) (angle p q t)))
:formula (not (implies
  (= (+ (angle p q s) (angle s q t)) (angle p q t))
  (< (angle p q s) (angle p q t))))

```

Data: all incidences (except on(u,N))

$\text{bet}(p, s, T)$

$\text{bet}(q, s, r)$

$\text{bet}(s, u, T)$

$p \neq q$

$\neg \text{on}(r, L)$

Implementation

Northrop is working on a rudimentary interface. There is an easy translation of Euclidean proofs:

When the user asserts a theorem: create the initial objects, assert hypotheses to the database, and remember the conclusion.

When the user types “let p be a point,” or some other construction rule: check prerequisites, create the new objects, and assert the relevant properties to the database.

When the user types “hence A ”: check A follows from the database, and if so, assert it explicitly.

For suppositional reasoning: push the state, assert the supposition, verify the conclusion, pop the state, and assert a conditional.

When the user types “QED,” check that the negation of the theorem’s conclusion is inconsistent.

Automated geometric reasoning

Approaches:

- Reduction to real closed fields, CAD: slow
- Wu's method: extremely powerful, but cannot handle order relations (like “between”).
- Area method: extremely powerful, produces readable proofs, complete for the class of “constructive linear theorems,” but once again cannot handle order.
- Synthetic methods.

(Google “automated geometry reasoning Chou Gao Zhang.” See also their “Machine proofs in geometry,” on the area method.)

Our approach falls into the last category.

Automated geometric reasoning

See Chou, Gao, Zhang, “A deductive database approach to automated geometry theorem proving and discovering.”

This provides an approach that is similar to ours, with much more sophisticated representations of diagrammatic information.

Our modest claims:

- We have a clean analysis of the type of reasoning that is used in books I–IV of the *Elements*.
- The analysis makes it easy to verify formal texts that are very close to proofs in the *Elements*.
- This provides a clear sense in which the *Elements* is more rigorous than commonly acknowledged, while raising interesting questions as to *why* diagrammatic inferences are so natural.

Conclusions

A broader lesson: an effective approach to formal verification is to use forward reasoning, building up contexts of data (in our case, “diagram information” and “metric information”) that can be handled in domain specific ways.

Another historical project: revisit the interplay of analytic and synthetic methods around Descartes.

More generally: try to find other focused domains where such an approach yields good results.