Constructive Geometrical Reasoning and Diagrams

John Mumma

In his writings on Hilbert's foundational work, Paul Bernays sometimes contrasts Euclid's geometrical method in the *Elements* with Hilbert's in *Foundations of Geometry*, identifying two features that distinguished the latter from the former (notably, in [6] and in the introduction to [3]). First, Hilbert's theory is *abstract*. Though the primitives of the theory have an intended geometric interpretation, a proof concerning these primitives depends only on logical form of the proof's premises. Euclid on the other hand presumes in proofs a geometric interpretation of his primitive notions. Second, Hilbert's theory has an *existential form*, while Euclid's is constructive. This second difference is nicely illustrated (as Bernays points out in [6] and [3]) by Hilbert's axiom I,1 and Euclid's first postulate. The former is

> For every two points A and B there exists a line a that contains A and B

while the latter is

> To draw a straight line from any point to any point.

With Hilbert's theory, the totality of its objects are assumed to exist in advance. The role of its axioms is to identify basic relations that hold between the objects of this totality—e.g. that for any two points in it there is a line in it such that the relation of containment obtains. On the other hand, Euclid's theory is *constructive*. The objects that are the subject of its proofs are constructed ones. Accordingly, it must include among its basic principles some that specify the primitive methods of construction available. Postulate 1, for example, identifies the construction of a line from two points as primitive.[1]

The first difference is one in geometrical reasoning, the second in geometrical ontology. My aim in this paper is to investigate the relation between them. Specifically, I explore the prospects of a sharp account of constructive geometry where a non-abstract understanding of geometrical terms is operative. It is well understood what an existential, abstract theory of geometry amounts to (the canonical paper is [14]). Further, an enormous amount of work has been done in clarifying the nature of constructive arithmetical

[1]The same contrast between Hilbert's axiom I,1 and Euclid's first postulate is made later by Mueller in [10](p. 14) and Chiraha in [4] (p. 8) to illustrate a point that in a general sense is the same as Bernays'.

reasoning. Little attention has been paid, however, to whether there exists a distinctly *geometrical* species of constructive reasoning—i.e. a species of constructive reasoning where the objects constructed are assumed to be geometrical ones.

My discussion is centered around **Eu**, a recent formal account of Euclid's diagrammatic reasoning ([11]). The paper has three parts. In the first, I motivate **Eu** as a candidate account by showing how modern, foundational theories of geometry fall short of being both non-abstract and constructive. In the second, I present the relevant details of **Eu** and argue that it offers a plausible account of Euclid's non-abstract geometrical reasoning. In the third part, I consider whether according to **Eu** Euclid's reasoning counts as constructive. Though there are general reasons to think so, there are challenges in working the claim out. The very aspect of the account that gives the proofs a distinctly geometrical character is naturally interpreted as requiring quantification over geometric objects. And so the account seems to place Euclid's proofs on the existential side of Bernays' existential/constructive divide. After describing this default interpretation, I comment, briefly, on what would be required to develop an alternate, constructive one.

Spatiality and abstract geometric theories

The philosophical problem of characterizing the objects of geometry is, of course, a difficult one. I take it to be uncontroversial, however, that a feature that any philosophical characterization must respect is their spatiality. If planes, lines and points are understood as geometric objects, then they have spatial properties, and realize spatial relations. Further, these spatial properties and relations are *intrinsic* to the geometric objects. Once a geometric object is given, it has spatial properties. Once two or more geometric objects are given, they realize a spatial relation. And so, if a theory of geometry is non-abstract—that is, if it assumes a geometric interpretation of its basic notions—its objects have spatiality from the start. No theoretical work is required to secure it for them.

Just the opposite is the case with Hilbert's abstract theory, as can be seen by considering the treatment of order relations within it. The theory has one primitive for such relations—a one dimensional betweenness relation. All other order relations are defined in terms of it. In particular, the two dimensional order relations between a point and a line within a plane are expressed in terms of betweenness. If we have any understanding of geometrical objects as spatial, we understand the fact that a line splits a plane into two halves, and a point can lie in either one. When one takes an abstract approach to geometry, however, this understanding is not available. Hilbert is thus obligated to express the position of a point with respect to the sides of a line in terms of the one dimensional order relation he has fixed as primitive.

The definition itself is worth looking at closely, for it also illustrates the existential assumptions of Hilbert's theory. The specific relation defined is that of two points p_1 and p_2 *lying on the same side of* a line l, given that the two points do not lie on the line. It is defined as

$$\forall\ points\ x\ \ B(p_1 x p_2) \rightarrow \neg\ l\ contains\ x$$

The negation of the relation, that of p_1 and p_2 *lying on different sides of* l is accordingly

$$\exists\ point\ x\ \ B(p_1 x p_2)\ \&\ \ l\ contains\ x$$

The idea behind the definition is that the position of two points with respect to a line is linked to the relative position of the segment joining the two points and the line. (See figure 1.) If the segment doesn't intersect the line,

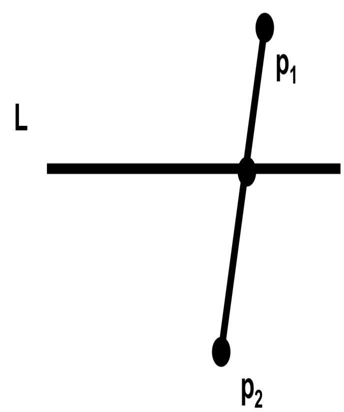


Figure 1: Configuration illustrating Hilbert's definition of opposite side

then the points lie on the same side. If it does, then the points lie on different sides.[2]

It is in the quantifiers of both definitions that Hilbert's existential assumptions reveal themselves. The definitions can only do what they are intended to do if their scope is taken to be all the points in the plane. Consider the definition of two points lying on opposite sides of a line as a biconditional.

$$OppSide(p_1, p_2, l) \quad \leftrightarrow \quad \exists\ point\ x \ \ B(p_1 x p_2)\ \&\ \ l\ contains\ x$$

From a constructive perspective, where what exists is what has been constructed, the biconditional does not seem to hold. Specifically, the left to right direction does not seem to be valid. The other direction is unproblematic, constructively. If a segment intersecting a line exists (i.e. has been constructed), then the endpoints of the segment must lie on opposite sides of the line. On the other hand, it is possible for a configuration to have two points lying on opposite sides of a line, without the segment joining the points, and the resulting intersection point, being constructed. Accordingly, if existence is interpreted as what has been constructed, then the left to right direction of the biconditional is false.

Hilbert of course does not interpret existence in this way. He takes the plane and all its parts to exist from the outset. In particular, if the segment between the points p_1 and p_2 intersects l, then that intersection point exists. No act of construction is required for the point to exist. And so there is no counterexample to the left to right direction of the biconditional. More generally, in assuming the existence of the plane and all its parts, Hilbert is free to regard any configuration C as contained in a larger, more complicated configuration $C*$, and define relations in C in terms of relations obtaining with the additional objects in $C*$. In contrast, from a constructive perspective, there are geometric relations in C that obtain whether or not a larger, more complicated $C*$ is constructed from C.

What these observations bring out is that the move to an abstract, existential theory of geometry flattens things out in a certain sense. There is

[2] As Hilbert's theory axiomatizes three dimensional geometry, another condition of these definitions is the containment of the objects l, p_1, p_2 and x within a single plane. With all the geometric examples discussed in the paper, this condition will be assumed. Further, in [5] Hilbert does not use a formal notation to make the quantifiers explicit, and the wording of the informal definitions and proofs suggests in fact a constructive standpoint. But this seems best understood as an expository quirk, as the existential form of the incidence axioms is explicit (though not formally so). For a discussion of the discrepancy between the incidence axioms and the informal exposition in [5], see [17].

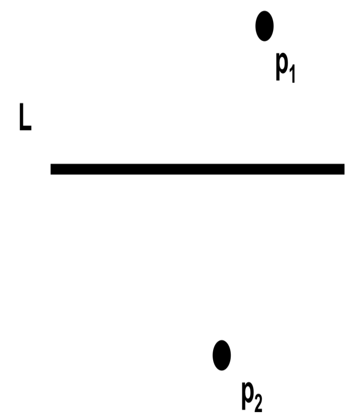


Figure 2: Opposite sidedness without a joining segment

no reason to regard the intrinsic spatiality of geometric objects as having a privileged, foundational status. In a non-abstract, constructive theory, on the other hand, the situation is different. A collection of geometric objects presupposes a construction sequence, where the existence of each object in the sequence is attributable to a construction step in the sequence. Any given collection can thus be understood to be preceded by a sequence of sub-collections, where the nth sub-collection is what has been constructed at the nth step. This builds in an asymmetry with respect to the final collection and its sub-collections. The relations obtaining in the final collection depend on the relations obtaining in the sub-collections in an irreversible way. Tracing the construction sequence back to the beginning, we see that the properties and relations on which everything else ultimately depends are those present in collections of one or two elements—i.e. the intrinsic spatial properties and relations of the geometric objects. When the totality of points, lines and planes are given at once, this dependency vanishes. As any collection C can be thought to exist simultaneously within a great many other larger configurations, there is no reason to regard the complex of relations obtaining in C as prior to the relations obtaining in the larger configurations.

Axiomatizations of geometry have been developed that can be thought to occupy a middle position. Though abstract like Hilbert's axiomatization, they develop geometry from a constructive perspective (in some sense of constructive). For an overview of this work, see [15] and [13]. It contains two strands. In one, initiated in [9] and pursued further by Pambuccian in series of articles (see [13] for the references), the theories are constructive in that their primitives consist entirely of operations. Their formal languages contain no relations (besides equality) or quantifiers. The axioms and theorems of elementary geometry are expressed as equations between different sequences of operations on points. In the other strand, found in [16] and [7], the languages of the theory contain relations and quantifiers, but these are interpreted in terms of a constructive theory of the continuum, which restricts the logic of the theory to intuitionistic logic. The theory developed in [2] combines both approaches. It is quantifier-free, contains relations, and is developed with intuitionistic logic.

It can be plausibly argued that each of these theories avoids, in some sense, the existential assumptions of Hilbert. Yet being abstract, they bypass the question of what it is for a single geometric object, or a geometric pair of objects, to be given.[3] Nothing is assumed of the theories' objects from the outset, and as a result the initial development of geometry within the theories mirrors that within Hilbert's. Definitions must be formulated and theorems proved to show that the objects of the theory have the spatial properties we expect geometric objects to have. Specifically, the order relations in a configuration C are sometimes defined in terms of more complicated configurations $C*$. The difference with Hilbert's definitions is that the quantifiers are replaced with constructive procedures that produce the additional objects of $C*$. For instance, in on p. 19 of [2], the opposite side relation is defined in terms of a betweenness relation B and the operations L and I, where L represents the operation of joining two points in a line, and I represents the operation of producing the intersection point between two lines:

$$Oppside(p_1, p_2, l) \leftrightarrow B(p_1, I(l, L(p_1, p_2)), p_2)$$

Though there is no quantification over points as in Hilbert's definition, it still demands the existence of the line joining p_1 and p_2 in order for the relation of

[3] von Plato explicitly acknowledges this in [16]. "This constructive geometry does not stipulate what the basic objects are, or how their basic relations are to be proved. In this sense it belongs to abstract mathematics, rather than to traditional constructive mathematics, where the aim has been to define once and for all the natural numbers and build all other mathematical structures on them" ([16], p. 92). The general question being explored here is whether an analogous constructive project with geometric objects is possible.

opposite-sidedness to enter into a proof. And this runs counter to a concrete, spatial understanding of lines and points, whereby we can talk of points lying on the same or opposite side of a line without having to construct the line between them.[4] There is nothing, of course, preventing the addition of a same-sidedness relation as primitive to an abstract constructive theory that already has betweenness as a primitive. But one would still be required to prove that the primitives relate within the theory according to their intended geometric interpretations.

In sum, with abstract theories of geometry (existential or constructive) spatiality is analyzed into a collection of axioms expressing particular fundamental spatial truths, where the axioms are composed of objects and relations explicitly designated as primitive. The acceptability of the axioms as a group depends, in part, on whether the full spatiality of the objects is logically recoverable from them. It depends, specifically, on whether non-primitive, but equally fundamental, spatial properties and relations can be defined, and whether fundamental truths about them and the primitives can be deduced. The aim of this section is to show how these initial steps seem unnecessary from a perspective where we take geometric objects to be spatial. Of course, this appearance does not mean that are in fact unnecessary in a mathematical sense. The question is whether a non-abstract approach to geometry can be defended as mathematically legitimate. That is, can we specify a mathematically precise form of geometric reasoning where a logically unanalyzed assumption of spatiality plays a part in what does and does not have to be proved?

The standardly presumed answer is no. Concrete notions of spatiality are thought to be too broad and open-ended to play a part in precise mathematical reasoning. They must, it is usually thought, be distilled into abstract, formal axioms if we are to have a clear picture of the structure of a geometric proof. The formal system **Eu** was developed to challenge this presumption. It specifies method where information from a representative diagram is integrated into proofs. By the method, geometric objects are assumed to be spatial from the outset in virtue of their representation by spatial, diagrammatic objects. Certain spatial properties and relations observable in diagrams are taken to be shared by the geometric objects represented. The intrinsic spatiality of the diagrams thus carries over to the geometric objects represented.

[4] Euclid, for instance, stipulates in the proof of I,7 that two points lie on the same side of the line *before* constructing the line joining them.

Moreover, the inferences where the spatiality of diagrams is used are tied to a constructive conception of geometric objects. And so the account promises a way to flesh out what constructive, non-abstract reasoning in geometry amounts to.

Reasoning with geometric diagrams

Any given geometric diagram manifests a wealth of observable spatial properties. This no doubt is a major reason behind the view that such objects cannot serve as instruments in careful mathematical argument. Careful mathematical reasoning demands control on the information being used at any stage in an argument, so that at no point is it unclear what justifies a step. The use of a diagrams as geometric representations might seem to make such a demand impossible to meet. The worry is that once diagrams with their rich array of spatial properties are used to represent geometric objects, the ability to isolate the standing of each claim in a geometric argument is compromised. Observing something in a diagram is too complicated, opaque a process to incorporate into mathematically controlled arguments.

This worry loses much of its force in the light of Ken Manders' investigations into Euclid's geometric proofs in [8]. His analysis reveals that diagrams serve a principled, theoretical role in Euclid's mathematics. Only a restricted range of a diagram's spatial properties are permitted to justify inferences for Euclid, and these self-imposed restrictions can be explained as serving the purpose of mathematical control. The aim of **Eu** is to build on Manders' insights, and precisely characterize the mathematical significance of Euclid's diagrams in a formal system of geometric proof.[5]

As a formal system of proof, **Eu** consists in a set of symbols, and rules for combining and transforming the symbols. Some of these rules serve to define what counts a well-formed expression, while others fix the permissible transformations on one or more well-formed expressions. The system of symbols and rules together is formal in that the rules are defined entirely in

[5]Another system E has been developed along the lines of **Eu**, and thus gives a formal picture of geometric inference very close to that of **Eu**. The key difference between them is the way diagrams are handled in each. **Eu** contains formal objects that model spatial diagrams. What corresponds to these objects in E are data lists that model instead *the information* directly extracted from spatial diagrams. For this reason I base my discussion on **Eu**. What recommends E is that it makes a meta-mathematical analysis of Euclid's diagrammatic method possible. For a full description of this analysis, see [1].

terms of the discrete syntactic form of the symbols. More precisely, the rules are recursive. There are effective ways to check that a sequence of symbols is well formed and to check that a transformation of well-formed expressions is permissible.

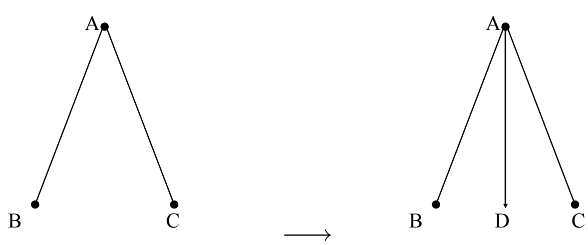
The notion of an **Eu** derivation is defined as a sequence of well-formed expressions that satisfy conditions stated in terms of the system's rules. **Eu** derivations are what model Euclid's proofs in **Eu**. They terminate with an expression of the form

$$\Delta_1, A_1 \longrightarrow \Delta_2, A_2$$

where Δ_1 and Δ_2 are diagrams, and A_1 and A_2 are sentences. The geometric claim this is stipulated to express is the following:

> ⋆*Given a configuration satisfying the non-metric positional relations depicted in Δ_1 and the metric relations expressed in A_1, then one can obtain a configuration satisfying the positional relations depicted by Δ_2 and metric relations specified by A_2.*

For an example, consider the **Eu** expression:



$BAD \equiv_{angle} DAC$

This is the **Eu** representation of proposition 9 in book I of the *Elements*, which states that given any angle, one can bisect it. (The slot for a metric condition A_1 in the antecedent is left blank because the proposition does not impose any metric conditions on the given angle.)

Manders' observation that only a restricted set of diagrammatic properties contribute in Euclidean proof is reflected in the role of an **Eu** diagram to depict *non-metric* positional relations. Though a diagram also possesses metric spatial properties, these do *not* figure in an **Eu** proof. The representation of these properties is achieved through the sentential symbols, just as it is done in the *Elements*. A close examination of the text shows that Euclid refrains from making any inferences that depend on a diagram's metric properties. It should not, on reflection, be surprising why this is. Geometric objects are ideal mathematical objects, and their ideality is linked to their metric properties. What distinguishes two line segments with respect to the metric property of length can be arbitrarily small, for instance. Physical diagrams cannot represent such exactness, and so in reasoning about it a different representational means are required. At the same time, however, it turns out that diagrams can serve as adequate representations of non-metric positional relations.

Eu demonstrates this by depicting geometric proof as running on two tracks: a diagrammatic one, and a sentential one. The role of the diagrammatic track is to record non-metric positional information of the figure, and to provide a means for inferring this kind of information about it. The role of the sentential one is to record metric information about the figure and provide a means for inferring this kind of information. The points, lines and circles of Euclids diagrams have formal analogues in **Eu** diagrams. The position of these formal analogues within the **Eu** diagram express information about the position of points to two dimensional objects (e.g. point p is on line l, point p is outside circle c) and the position of two dimensional objects to one another (e.g. lines l_1 and l_2 intersect, line l does not intersect circle c). **Eu** sentences express relations between magnitudes in **Eu** diagrams.[6] To understand how it is that an **Eu** diagram can be responsible for only non-metric positional relations, consider the **Eu** diagram in figure 3. expresses the relation: A and B lie on the same side of line segment CD. It expresses nothing, however, about the distances between the four points in the diagram. Consequently, when an **Eu** diagram is paired with a sentence expressing metric information there is nothing which requires that the diagram model this information. For example, in **Eu** a diagram looking like the diagram of figure 4 can be paired with a sentence which asserts that the magnitudes BC and EF are equal. As the diagram is responsible for only non-metric positional information, it need not conform to the metric assumptions of the proof. The diagrams that appear with elementary arguments in geometry standardly do conform to the stated metric assumptions, so it may seem that the **Eu** constraints

[6]For some of the details of **Eu**'s syntax, see the appendix of [11] and for all of the details, see sections 1.2. and 1.3 of [12]. The exposition in both uses Manders' terminology 'exact' and 'co-exact' to describe what is referred to here as 'metric' and 'non-metric positional.'

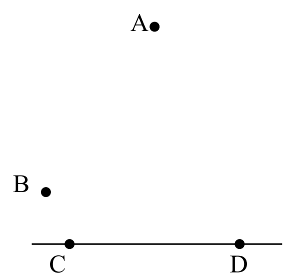


Figure 3: Eu diagram expressing *only* that A and B lie on the same side of CD.

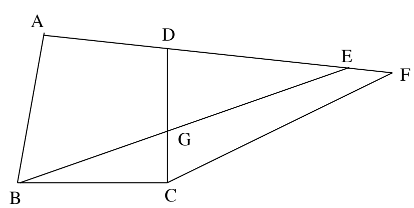


Figure 4: Inexact but acceptable **Eu** diagram

on diagrams are too loose. One can, however, produce with such symbols derivations which match Euclid's reasoning in the *Elements* closely. See [11] and [12].

Further, the representation of non-metric positional relations via diagrams is enough to build spatiality into geometric objects at the ground level. The range of possible order relations among geometric objects is directly reflected in the range of possible order relations between the elements of a diagram. As a result, *any* order distinction with respect to a geometric configuration can be recognized in a proof without prior definitions being given and prior theorems proved. The distinctions already apply to the very representations used to reason about them. The way to express, for the sake of an **Eu** proof, that two points lie on the same side of a line (as opposed to opposite sides of a line) is to produce a diagram where two diagrammatic points lie on the same side of a diagrammatic line. **Eu** thus respects the intrinsic spatiality of geometric objects in a way that modern, abstract theories do not.

The reasoning carried out with diagrams in **Eu** is conditioned by two features of Euclid's proofs: their generality, and their use of auxiliary geometric objects. Their generality is reflected by $\star$ above. The proof of a Euclidean proposition, if correct, establishes something not for just one geometric configuration, but for *all* configurations satisfying the conditions of the proposition. The second feature is present in proofs that contain a construction stage. In this stage, the initial geometric configuration of a proposition is augmented with additional geometric objects. With the help of these objects the desired conclusion is then shown in a subsequent demonstration stage. In **Eu**, the construction stage is modeled as the augmentation of an initial representative diagram with additional diagrammatic objects and, possibly, the introduction of metric assertions. The demonstration stage is then modeled as a sequence of inferences that are drawn from the symbols produced in the construction stage, where everything inferred holds of all configurations within the scope of the proposition.

The diagrammatic inferences in this sequence concern the newly constructed objects of the proof. In particular, they concern the non-metric positional relations of these objects to the objects of the initial configuration, or to one another. As diagrams are spatial, a host of new positional relations fall out with every augmentation. Every added diagrammatic point has a position with respect to the pre-existing diagrammatic lines and circles. Every added diagrammatic line or circle introduces a new partition of the plane in which the diagram lies, and all the pre-existing objects of the diagram have some determinate relation to the partition. The resulting collection of manifest relations cannot, however, be simply read off from the diagram and used in a proof. And this is because the diagram is intended to represent a range of geometric configurations. Nothing in its bare appearance separates relations that are representative from those that depend on the initial diagram's particular features. A process of reasoning is required to make the separation.

In the reasoning process prescribed by **Eu**, one must focus one's attention on each construction step. A construction step (e.g. the joining of two points in a segment) produces an object y (e.g. a segment) from a tuple of objects $\vec{x}$ (e.g. two points). If y and the objects of $\vec{x}$ are so related, say that y *directly depends* on the objects of $\vec{x}$. The relation is asymmetrical, and so naturally understood as an ordering relation. We can extend it to all objects of the augmented diagram by taking its transitive closure, and so obtain a partial ordering $\triangleright$ that records the dependencies of the diagrams objects. To describe how this partial ordering enters into the interpretation of the diagram, it is useful to introduce one more term. Define the relation *is linked* as the symmetrical version of 'directly depends'—i.e. x and y are linked if x directly depends on y or y directly depends on x.

Accordingly, the problem of isolating the general in an augmented diagram comes down to determining which of the unlinked objects in the diagram display general relations. The relations exhibited by linked objects are automatically general. It is built into the concept of a construction step that the constructed object y has a certain positional relation to the objects $\vec{x}$ it was constructed from. If a line segment l is constructed from points p_1 and p_2, for example, then it is immediate that l passes through p_1 and p_2. Also, if a circle c is constructed from center p and radius r, it is immediate that c contains p and r.[7] Generality cannot be said to be built into the relations exhibited by unlinked objects in a similar way, however. And so, by the lights of **Eu**, it is indeterminate whether these relations depend on the particular features of the diagram or not. **Eu** requires further reasoning to establish that they are not.

This reasoning follows the paths laid out by the partial ordering $\triangleright$. The objects at the bottom of the poset are those making up the proofs initial diagram, and so exhibit relations understood as general from the outset. The reasoning works its way up $\triangleright$, establishing the generality of relations between objects above the bottom level.

The basic inferences constituting this reasoning have a general form, embodied in the following schema (called *ADI* for augmented diagram inference):

$$ADI \ \frac{R_1(\vec{f},\vec{x}) \quad C(\vec{x},y)}{R_2(\vec{f},y)}$$

$R_2(\vec{f},y)$ in the schema denotes the positional relation in the diagram inferred

[7]These two constructions are postulated by Euclid, but a construction step does not have to be an instance of one of Euclid's postulates for it to induce the relation of direct dependence. A construction which is the application of a proven problem can also induce the relation. So, for instance, if one applies proposition I,1 to a segment l to construct a triangle on that segment, the elements of the triangle distinct from l are directly dependent on l, and thus linked. The rules of **Eu** are sound in a geometric sense, and this guarantees the generality of the positional relations in proven constructions.

as general. $C(\vec{x}, y)$ denotes a basic construction step, which produces y from the tuple $\vec{x}$, and $R_1(\vec{f}, \vec{x})$ denotes a collection of positional relations established as general between objects $\vec{f}$ and $\vec{x}$. The tuple $\vec{f}$ is best understood as a frame of reference. $R_1(\vec{f}, \vec{x})$ is the condition that the objects of $\vec{x}$ have a certain position with respect to $\vec{f}$. This position along with the nature of the construction $C(\vec{x}, y)$ licenses the inference $R_2(\vec{f}, y)$—i.e. y has a certain position with respect to the frame of reference $\vec{f}$. What's behind the inference, specifically, is the recognition of a positional invariant. $R_1(\vec{f}, \vec{x})$ does not fix the *exact* position of $\vec{x}$ with respect to $\vec{f}$ but specifies a range of positions over which $\vec{x}$ can vary. This limitation in $\vec{x}$'s positional relationship to $\vec{f}$ forces a positional relationship of y to $\vec{f}$.

Suppose, for example, that f is a single circle, $C((p_1, p_2), y)$ is the construction of ray y through p_2 from endpoint p_1, and $R_1(f, (p_1, p_2))$ fixes the position of p_1 inside f. This forces the intersection of y with f, which we can denote as $R_2(f, y)$. No matter where p_1 sits inside f, y has to intersect it. How the inference is codified diagrammatically in **Eu** is shown in the upper left corner of figure 5, which contains three other examples of *ADI* inferences. In each, the diagram on the top left depicts $R_1(\vec{f}, \vec{x})$, the diagram on the top right $C(\vec{x}, y)$ and the diagram below the line depicts the invariant relation $R_2(\vec{f}, y)$. Every example is an inference that occurs in the *Elements* by the analysis of **Eu**. The *ADI* inference of the top left inference occurs in the **Eu** analysis of I,2, that of the top right in the analysis of I,1, that of the bottom left in the analysis of I,6 and 1,16, and that of bottom right in the analysis of I,10. It is worth noting that the legitimacy and fundamental nature of inferences of this form is attested by the fact that they appear as axioms in modern axiomatizations of elementary geometry. Pasch's axiom (so named because Pasch first formulated it) and axiom $A13'$ in Tarski's theory $\mathcal{E}_2''$ [14] are examples.[8]

That a diagrammatic inference has an *AD* form does not guarantee, of

[8]In words, the Pasch axiom is 'any line constructed from a point on one side of the triangle intersects one of the triangles two other sides.' It is thus a proposition of the form

$$\forall \vec{f}, x, y \;\; (R_1(\vec{f}, x) \;\&\; C(x, y)) \longrightarrow R_2(\vec{f}, y)$$

where $\vec{f}$ is a triangle, R_1 expresses that the point x lies on a side of the triangle, $C(x, y)$ expresses that the line y is constructed from x, and R_2 expresses that the line y intersects one of the triangles other two sides. Tarski's $A13'$ is a proposition of the same form corresponding to the upper left inference in figure 5.

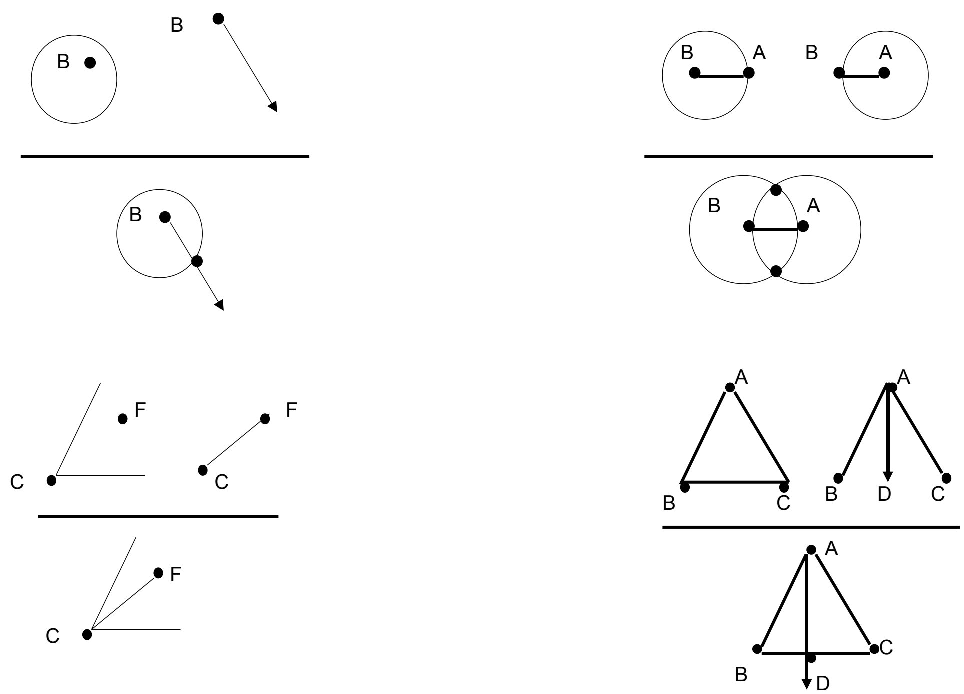


Figure 5: Some ADI inferences

course, that the inference is sound. There are indefinitely many ways to define a frame of reference $\vec{f}$ in terms of line segments and circles as well as indefinitely many possible construction steps $C(\vec{x}, y)$, and so indefinitely many ways to position a construction $C(\vec{x}, y)$ with respect to a frame of reference $\vec{f}$. Some of these force a positional invariant, others do not. Being a fixed formal system, **Eu** distinguishes a group of those that force an invariant as fundamental inference rules, and requires that the rest be derived. It thus possesses an artificiality akin to Hilbert's definition of same-sidedness in terms of betweenness. In contrast to Hilbert's theory, however, one can read from **Eu** a quasi-formal method of geometric reasoning where this artificiality is removed. Working according to such a method, one would still obligated to link unlinked elements to use their manifest positional relations in a proof. But one would free to make the link via AD inferences that have not been connected to a pre-specified list. One would be free, in particular, to call upon one's understanding of the geometric concepts the diagram is thought to represent. This comes down to understanding how $\vec{x}$ can vary with respect to $\vec{f}$ according to $R_1(\vec{f}, \vec{x})$, and the consequences of this variation with respect to the position of y to $\vec{f}$. From this understanding one is permitted to infer a relation $R_2(\vec{f}, y)$.

Eu thus provides a quasi-formal picture of geometric construction and proof where a non-abstract, spatial understanding of geometric objects is operative. In contrast to abstract geometric theories, the spatial properties of an initial configuration have a foundational status in an **Eu** proof. This is reflected directly in the $\triangleright$ ordering. The ordering is induced by a sequence of constructive acts carried out against a spatial background. Each act of construction satisfies spatial conditions, which when coupled with the spatial character of the object constructed, licenses certain spatial inferences (i.e. the *AD* inferences). As diagrams are what carry these inferences, the reasoning goes beyond the mere logical form of the geometric premises and conclusion. The diagrams provide a means, as it were, to apprehend their spatial form. And so, the **Eu** account falls opposite HIlbert's theory with respect to Bernays' abstract/non-abstract distinction.

The question of the next and concluding section is where the account sits with respect to Bernays' second, ontological characterization of the difference between Hilbert and Euclid. Though geometric constructions are central to the **Eu** proof method, it is not clear the method avoids the presumption of an infinite totality of geometrical objects. The question comes down to how an **Eu** derivation ought to be analyzed as proving a general geometric proposition. In understanding it as proving statements about a range of geometric configurations, realizing a continuum of possible (exact) positional relations, does one presume the existence of an infinite totality of geometric objects?

Conclusion: generality and ontology in elementary geometry

If one steps back and temporarily puts aside the question, the idea that the diagrammatic proof method of **Eu** is constructive is attractive. The existential/constructive contrast in geometry lines up naturally with another one: geometry as a theory of space, and geometry as a theory of spatial figures. Existential geometry corresponds to the former, while constructive geometry corresponds the latter. Bernays more or less makes such a correspondence in [3], as it provides a nice historical illustration of the general development of modern mathematics away from concrete constructions to abstract, existential axiomatics.[9] Geometry before the 19th century concerned constructed spatial figures. Geometry after the 19th century possessed abstract axiomatic characterizations of entire spaces. It became possible to talk of euclidean space, and distinguish it from others.

Eu could be thought to round this story out by providing a picture of what constructive geometric reasoning predominant before the upheavals of the 19th century amounts to. Its formal structure as a proof system constrains geometrical discourse to what can be represented in an **Eu** diagram. The objects reasoned about are always finite, both in being bounded—no diagram has an infinite spatial extension—and being composed of fixed, finite number of objects—no diagrammatic construction sequence proceeds beyond ω. Finiteness is not compatible with lines and circles being understood as sets of points, of course, but as Bernays points out in section 1 of part II of [3], this understanding is not forced upon us. Equally viable is a conception where a line, for instance, is an unitary whole that marks out possible positions for points.

Nevertheless, if **Eu** derivations are to be understood as proofs, what is being proved are always general propositions. The antecedent of a derived **Eu** expression is not intended to designate a singular geometric configuration, but to specify conditions satisfied by a range of configurations. The logical form of the entire expression is

$$\forall \vec{x} \exists \vec{y} \quad \phi(\vec{x}) \rightarrow \psi(\vec{x}, \vec{y})$$

where ϕ stands for the initial conditions of the geometric claim, and ψ the properties to be proved. If a logical analysis is to be carried through to individual steps in **Eu** proofs, other propositions with quantifiers appear. The inferences of the demonstration stage, including the *AD* inferences (see footnote 8), are universal statements when expressed as propositions. In the construction stage, most steps produce unique geometric objects from given ones and so can be represented logically by functions. Yet one kind does not:

[9]Bernays remarks on the historical development of geometry from the study of figures to the study of spaces at the beginning of section 1 of part I in [3]. The connection between the figure/space distinction and the constructive/existential distinction comes out in section 1 of part II, when Bernays argues that the infinite is not given in geometric intuition.

the free choice of a point satisfying non-metric positional conditions. Such points have an indefinite character within proofs. Their precise identity is not fixed relative to other given objects in the configuration. The natural logical representation for what licenses their introduction are thus existential statements, asserting the existence of a point satisfying certain non-metric positional conditions.[10] And so, though the geometric reasoning in **Eu** is always performed with a particular finite diagram, it still seems to presuppose a domain of geometric objects, i.e. the domain over which the quantifiers of these propositions range.

The upshot seems to be that Euclid's figure based reasoning is existential, despite Bernays' constructivist portrayal of it. At the very least, what can be safely asserted is that Euclid's figure based reasoning differs from the traditional conception of constructive reasoning in mathematics. This conception is essentially arithmetical (see the quote from von Plato in footnote 3), and is based on constructions that produce *particular* numbers. The fundamental propositions of a constructive theory of arithmetic are identities between such constructions. These are the particular facts on which arithmetical theories advanced as constructive are built. According to **Eu**, there are no analogous identities at the foundation of Euclid's geometry. Constructions have a general scope from the start. Identities can be asserted between different geometric constructions, but the identities do not concern particular geometric objects but a range of geometric objects.

The theorem of elementary geometry that the three angle bisectors of a triangle ABC intersect in a point can be understood, for instance, as equating two constructions: the intersection of bisectors of the angles ABC and BCA, and the intersection of the bisectors of the angles BCA and CAB. Yet the triangle ABC is not one individual triangle. Nothing about its position or orientation is specified, nor is anything specified about the relative magnitude of its sides and angles. All these can vary continuously, and the theorem still applies. The identity thus cannot be thought to be analogous to a basic arithmetical identity like

$$1 + 2 = 2 + 1$$

[10]Though Euclid does not introduce indefinite points into his proofs often, he does do it. For examples, see propositions 5, 9, 11 and 12 of book I, and proposition 1 of book III.) The introduction of such objects in a proof can be thought to be dual to the AD inferences. The latter add information about objects satisfying non metric positional conditions, while the former add objects satisfying non metric positional conditions into proofs.

It corresponds rather to arithmetical *laws* like

$$\forall x, y \quad x + y = y + x$$

The deep difference here is that geometric constructions construct spatial objects within a spatial continuum, while arithmetical constructions are more abstract and require no special medium. As a result, the setting for the initial step of a geometric construction always can differ in geometrically significant ways, while with basic arithmetical constructions any variation in the setting is arithmetically irrelevant. Specifically, a geometrical construction invariably involves parameters that can vary continuously. In the example of the bisector intersection theorem, one such parameter is the position of the point C with respect to the segment AB. This determines the relative magnitudes of its sides and angles, and is geometrically relevant for the construction. Variation in the parameter leads to variation in the geometric properties of the configuration constructed. The task of the constructive geometer, in this case and in general, is to isolate those properties in the configuration that obtain for all values of the construction's free parameters—in other words, to find the invariants of the construction. No analogous variation exists with the basic constructions of arithmetic. The successor operation can be understood to act on a variety of different objects, but the specific character of the object has no arithmetical consequences with respect to the object constructed. Seven strokes have the same arithmetical properties as seven dots.[11]

Eu provides an account of how the constructive geometer's task can be carried out systematically with diagrams. The geometer scans the diagram in terms of its construction, and determines whether certain AD inferences are justified. The account does not, however, provide any immediate guidance regarding the ontological status of elementary geometry. The question comes down to the notion of continuous variation underlying the AD inferences. How is it be analyzed? What is involved, for instance, in the judgment that for *any* ray emanating *anywhere* inside a circle, the ray intersects the circle?

[11] The observation here echoses Kant's argument in the *Critique* that arithmetic possesses no general propositions (i.e. 'axioms') that are also synthetic: "If I say 'With three lines, two of which taken together are greater than the third, a triangle can be drawn' then I have here the mere function of the productive imagination, which draws the lines greater or smaller, thus allowing them to abut at any arbitrary angle. The number 7, on the contrary, is possible in only a single way, and likewise the number 12, which is generated through the synthesis of the former with 5."(A164/B205)

An existential interpretation of Euclid's elementary geometry provides a straightforward and quick answer. The full continuum of points in the circle exist, as do all the possible rays from each point. The judgment is then correct insofar each of these rays shares a point with the circle. An alternative, constructive interpretation of the inference is not immediately apparent. At the same time, however, it is not immediately apparent that one cannot be developed. What would be required is an in-depth philosophical analysis of geometric continuity and its role in our mathematical reasoning. The analogous notion in arithmetic is induction, as it is what underlies general arithmetical statements like the commutative law of addition above. There has been a great deal of philosophical work investigating the justification and ontological significance of various forms of arithmetical induction. If a reading of Euclid as a constructive *and* geometrical theory is possible, similar investigations into geometric continuity (which do not merely reduce it to logical and/or arithmetical characterizations) are called for.

References

[1] J. Avigad, E. Dean, and J. Mumma. A formal system for euclid's *Elements*. *Review of Symbolic Logic*, 2:700–768, 2009.

[2] M. Beeson. Constructive geometry. to appear in Arai, T. (ed) *Proceedings of the Tenth Asian Logic Colloquium, Kobe, Japan*. Available at www.michaelbeeson.com/research/papers/pubs.html.

[3] P. Bernays. Die philosophie der mathematik und die hilbertsche beweistheorie. In *Abhandlugen zur Philosophie der Mathematik*, pages 17–61. Wissenschaftliche Buchgesellschaft, Darmstadt.

[4] C. Chihara. *A Structuralist Account of Mathematics*. Clarendon Press, Oxford, 2007.

[5] D. Hilbert. *Foundations of Geometry*. Open Court Publishing Company, La Salle, Illinois, 1971.

[6] D. Hilbert and P. Bernays. *Grundlagen der Mathematik*. Berlin, 1935.

[7] M. Lombard and R. Vesley. A common axiom set for classical and intuitionistic plane geometry. *Annals of Pure and Applied Logic*, 95:229–255, 1998.

[8] K. Manders. The euclidean diagram. In *Philosophy of Mathematical Practice (Mancosu, editor)*. Clarendon Press, Oxford, 2008.

[9] N. Moler and P. Suppes. Quantifier free axioms for constructive plane geometry. *Compositio Mathematica*, 20:143–152, 1968.

[10] I. Mueller. *Philosophy of mathematics and deductive structure in Euclid's* Elements. Dover, 1981.

[11] J. Mumma. Proofs, pictures, and euclid. to appear in *Synthese*.

[12] J. Mumma. *Intuition Formalized: Ancient and Modern Methods of Proof in Elementary Euclidean Geometry*. PhD dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2006.

[13] V. Pambuccian. Axiomatizing geometric constructions. *Journal of Applied Logic*, 6:24–46, 2008.

[14] A. Tarski. What is elementary geometry? In *The Axiomatic Method, with Special Reference to Geometry and Physics (Henkin, Suppes, Tarksi,)*. North Holland Publishing Company, Amsterdam, 1959.

[15] R. Vesley. Constructivity in geometry. *History and Philosophy of Logic*, 20:291–294, 2000.

[16] J. von Plato. The axioms of constructive geometry. *Annals of Pure and Applied Logic*, 76:169–200, 1995.

[17] J. von Plato. Formalization of hilbert's geometry of incidence and parallelism. *Synthese*, 110:127–141, 1997.