Construction Theorems and Constructive Proofs in Geometry

John B. Burgess

Abstract: Given Tarski's version of Euclidean straightedge and compass geometry, it is shown how to express construction theorems, and shown that for any purely existential theorem there is a construction theorem implying it. Some related results and open questions are then briefly described.

Keywords: Euclidean Geometry, Straightedge and compass, Construction problems, Existence theorems

1. Introduction

I will be concerned with comparing and contrasting three types of mathematical assertions, illustrated by these:

(1) Existence claim:
There exists a regular heptadecagon.

(2) Constructibility claim:
It is possible to construct a regular heptadecagon.

(3) Construction claim: It is possible to construct a regular heptadecagon in the following way ...

where the ellipsis in the last item would be completed with the specification of Gauss's construction or some other (as in Weisstein, 2021).

The kind of constructions meant here are those familiar from the early books of Euclid's *Elements*, traditionally called *straightedge and compass* constructions. What would have to be meant are an *unmarked* straightedge, not usable as a ruler, and a *collapsing* compass, not usable as dividers, since a ruler or dividers would make it a trivial matter to transfer a length, while Euclid takes doing so to require the sort of substantive material one finds in his Proposition 2. Actually, what Euclid assumes is that a line can be drawn through two points A and B in the plane (in two steps, first joining the points to form the segment AB as provided for by Postulate 1, then extending the segment indefinitely in a straight line at either end, as provided for by Postulate 2), and also that a circle can be drawn of given center and given radius (as per Postulate 3). And he does not specify any method or tools for doing such things; in particular, he does not mention the use of straightedge and compass. Let us nonetheless for convenience retain the traditional label for the class of constructions in question.[1]

John Burgess, Princeton University, USA

Journal for the Philosophy of Mathematics, Vol. 1, 23-41, 2024 | ISSN 3035-1863 | DOI: 10.36253/jpm-2932

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.
Competing Interests: The Author(s) declare(s) no conflict of interest.

Firenze University Press | https://riviste.fupress.net/index.php/jpm

[1] But it may be not inappropriate to recall parenthetically that there are other ways to draw lines and circles, some antedating the oldest surviving examples of compasses, which come from Roman times. Proclus describes geometry as arising out of the practice of Egyptian surveyors, with which Euclid as a resident of Alexandria might be expected to have been familiar. These surveyors were called "rope stretchers" and are depicted in wall paintings as carrying a long rope, which stretched taut can be used to trace a straight line on the ground. It is no coincidence that our word "straight" is etymologically linked to "stretched' and that "line," which is still the nautical term for "rope," is etymologically linked to "linen," flax being one of the materials out of which ropes were produced. A method something like this, involving a taut cord tied at both ends to stakes, which I learned as a child from my grandfather, a landscaper, is still used today in gardening, and a web search on the key words "garden" and "straight line" will turn up several videos giving a demonstration. A taut string could be used to draw a straight line on paper or papyrus or parchment in an analogous way. A string can also be used to draw a circle. It can even be used to draw an ellipse, though Euclid makes no provision in the *Elements* for ellipses and other conic sections.

Many propositions of Book I, beginning with the very first, are "problems," whose solutions end with words amounting to "which was to be done" or QEF, in contrast to "theorems," whose demonstrations end with "which was to be shown" or QED. Both kinds of propositions have in Euclid a very stylized form of exposition, according an analysis of Proclus (see Netz 1999) consisting of the same sequence of a half-dozen parts: *protasis, ekthesis, diorismos, kataskeve, apodeixis, symperasma*. In a problem, the *kataskeve* does what is to be done, and the *apodeixis* shows that it has successfully done it. In a theorem, the *kataskeve* introduces auxiliary points, lines, circles, or whatever, and the *apodeixis* uses them to show what was to be shown. The term "construction" gets used in two ways in discussing these matters, first as a term for problems as opposed to theorems, second as a translation of *kataskeve*.

Philosophers of mathematics have often contrasted the ancient style of axiomatic geometry represented by Euclid with the modern style represented by David Hilbert (1902). Paul Bernays (1964, p. 275) explicitly draws the contrast as one between statements of type (2) and statements of type (1):

> Euclid postulates: One can join two points by a straight line; Hilbert states the axiom: Given any two points, there exists a straight line on which both are situated.

Thomas Heath, however, in his classic translation (1968), renders Euclid's formulations as infinitival phrases ("to join two points…") rather than complete sentences ("one can join two points…") of type (2). But probably it would not matter much to Bernays if his formulation of Euclid had to be confessed to be not quite accurate, since a closer look at the context shows that Bernays is much less concerned with differences between Hilbert and Euclid (who indeed is mentioned only briefly very near the beginning of the paper) than with differences between Hilbert and twentieth-century "constructivists."

Chief among these was L.E.J. Brouwer, the founder of mathematical intuitionism and Hilbert's main adversary in the *Grundlagenstreit* or foundational dispute of the years between the world wars, which was just beginning to wind down in 1934 when Bernays produced the original French version of the paper just quoted. And Euclid and Brouwer differ from Hilbert in different ways.

2. Two Kinds of Constructivity in Geometry

Where Hilbert would have existence theorems, Euclid would have construction problems, while Brouwer would still want existence theorems, but would impose a requirement rejected by Hilbert, that a proof of a theorem to the effect that there exists a mathematical object satisfying a given condition not be accepted unless it is "constructive" in the sense that there is implicit in it a method of specifying a particular example of such an object (as for instance Euclid's proof in the *Elements*, Book IX, Proposition 20 that for any given number n of primes there is a further prime, has implicit in it the method for finding such an additional prime, namely, taking the product of the given primes, adding one to obtain a number m, and checking all numbers up through m until a prime dividing m is found.)

Brouwer *identified* mathematical existence with the possibility of being constructed (or in his more extreme formulations, with the actuality of having been constructed). But in this Brouwer avows influence not from Euclid but from Kant's view that mathematical proof involves "constructions in intuition" (or as Brouwerians sometimes put it, "mental mathematical construction"), though the Kantian view itself was clearly influenced by the role of construction in ancient mathematics, so there would be a Euclidean influence on Brouwer after all, at least at one remove.

Their requirement of constructivity in proofs makes it impossible for intuitionists to accept the classical logical laws of the *excluded middle*, $p \vee \neg p$, or equivalently the law of *double negation* $\neg\neg p \rightarrow p$, as used in proofs by contradiction or *reductio ad absurdum*. Brouwer's disciple Arend Heyting presented a formal version of the principles intuitionists *do* accept, and these two are conspicuously absent, and have to be, because using them one can easily produce existence proofs where no specific example of the sort of thing claimed to exist is provided even implicitly.

What is perhaps the simplest case of this phenomenon is shown in the classical proof, intuitionistically unacceptable, of the following:

(4) $\exists x \forall y(\Phi(y) \rightarrow \Phi(x))$

"There is something that Φs if anything does." For excluded middle tells us that either there exists something that Φs or there doesn't. If there does, any such thing may be taken for x in (4), and the conditional will be true because its consequent is true. If there does not, then any x at all may be used in (4), and the conditional will be true because its antecedent is false. This clearly does not give us a specific example of an x of the kind asserted to exist by (4), unless we already either know an example of something that Φs or know that there isn't any.

It would be anachronistic to say that Euclid reasons by "classical" logic in the sense of the logic expounded in orthodox present-day textbooks, but he certainly does *not* conform to the constructivist restrictions of intuitionistic logic. Notoriously he uses *passim* the method of proof by *reductio* (see for instance Book I, Proposition 29). And according to the early modern commentator Christoph Schlüssel, he uses the equally intuitionistically unnacceptable law $(\neg p \rightarrow p) \rightarrow p$, sometimes called the *consequentia mirabilis*, but dubbed by Jan Łukasiewicz the *law of Clavius* ("Clavius" being Schlüssel's Latinization of his Germanic family name, which like the Latin *clavis* means *key*) and perhaps better known under that label. So we must distinguish Euclid's emphasis on *construction problems* from Brouwer's insistence on *constructive proofs*, though the full significance of the difference between the two will only gradually become clearer in what follows.

It appears to have been Bernays' paper that first introduced something like the (historically dubious) use of "platonism" current in philosophy of mathematics during the last half-century or more, during which period there has been a vigorous debate between so-called platonists and so-called nominalists over the existence of abstract entities, and specifically mathematical objects such as numbers or sets. Allusions to Euclidean constructions have played an occasional role in this debate.

Notably, Charles Chihara (1973), one of the earliest of a group of writers who have attempted to reconstruct or reconstrue mathematics nominalistically, pursues the following sort of strategy. First, statements about mathematical objects such as numbers are replaced by statements about linguistic expressions, mathematical notations such as numerals. Second, assertions of the actual existence of linguistic expressions as abstract types are replaced by assertions of the possible existence of linguistic expressions as concrete tokens. Third, the modal notion of "possible existence" in this context, which some have criticized as unacceptably "metaphysical," is explained as potential *physical* "constructibility," which Chihara claims is a notion that should be familiar from ancient mathematics and considered philosophically respectable.

And in this last connection he quotes with approval an expression of a similar sentiment from his colleague Ernest Adams (1973, p. 406), who alludes to Euclid thus:

> Euclidean geometry cannot be criticized for lack of rigor simply on the grounds of its modal formulation, whatever other faults it may have in this respect, since the logical laws to which the constructibility quantifier conforms are quite clear.

It is clear enough, though, from the writings of the contemporary and recent philosophical thinkers mentioned so far—Adams, Bernays, Brouwer, Chihara, Heyting, and the list could be extended—that even if all are inspired in some way and to some degree by Euclid's *Elements*, none is primarily concerned with arguing over the correct interpretation of that ancient text. But apart from appropriations for other philosophical purposes, the correct interpretation of Euclid on construction *has* been a topic of interest for its own sake, and a subject of discussion among scholars of ancient Greek philosophy and historians of ancient Greek mathematics.

Recently Silvia De Toffoli and I conducted a joint graduate seminar on mathematical rigor in theory and practice, and among our guest speakers we were fortunate enough to have Benjamin Morison, who gave an account of his work with Jonathan Beere (unpublished manuscript) in precisely this area. The present note is a kind of scholium to the Beere-Morison paper, describing some relevant logical facts about formalized systems of geometry, in a way that I hope will be readable by and informative to students of classical philosophy and ancient mathematics who have a modicum of knowledge of modern logic but are not specialists.

The main thesis of the joint authors concerns the *purpose* of construction problems. (They also offer the opinion that the infinitival phrases "to do this or that" should be taken to have a kind of imperatival or jussive

force; but philological issues are beyond me.) Their view is that, just as theorems and their proofs aim to impart an especially solid kind of knowledge-*that* something is the case, *episteme* or *scientia*, so construction problems and their solutions aim to impart an especially solid kind of knowledge-*how* to do something. It is when one tries to express this knowledge-how as a kind of knowledge-that that one may arrive at something like a construction theorem of type (3), as contrasted with a constructibility claim of type (2).

Constructibility claims may still play a role, though a limited one, for those whose primary interest is in construction claims. How is perhaps best brought out by comparison with the views of finitists on arithmetic. For finitists, existence statements in number theory of the type "condition $\Phi(n)$ holds for some natural number n" are not really meaningful or *inhaltich*. Nonetheless, finitists allow that such a form of words may be used as a *partial communication* of meaningful results of the type "condition $\Phi(n)$ holds for the following natural number n..." where would follow the specification of a relevant n. In the same way, something like (2) may be admitted as a *partial communication* of something like (3).

As Beere and Morison make clear, though existence and constructibility assumptions or assertions are not entirely absent from the *Elements*, they are rare. By contrast, construction problems are ubiquitous in the geometrical books, which famously culminate in Book XIII with the construction of the regular polyhedra or Platonic solids. The situation is different in present-day mainstream mathematics. There questions of existence are prior, and it is only when one is answered in the affirmative that a question of constructibility then arises, or rather a variety of them for various kinds of constructibility (of which straightedge and compass constructibility in plane geometry is only the best known). Similarly with numerical functions: Existence comes first, computability in this or that sense (say recursive or primitive recursive or polynomial-time) being a distinct and subsequent issue or series of issues.

Moreover, though modal idioms such as "it is possible to..." or "one can..." or *-ible* and *-able* suffixes do occur all over present-day mathematics, including the present note, in formal contexts such locutions get explained away in terms of existence statements, rather than the other way around. Thus constructib*ility* and comput*ability* are defined in terms of the existence of programs for constructing or computing. This circumstance would affect the understanding of (2) and (3), turning them into "There is a program for doing such-and-such," and "The following is a program for doing such-and-such...," or something of the sort.

3. Tarskian Rigor

But the most important difference of present-day from ancient mathematics is that mathematicians now hold themselves to a higher standard of rigor. They have done so since the later nineteenth century when Frege opened the body of his epochal *Grundlagen der Arithmetik* with the following much-cited remark, which I quote in the English translation of J. L. Austin (see Frege, 1953, p.1):

> After deserting for a time the old Euclidean standards of rigor, mathematics is now returnng to them, and even making efforts to go beyond them.

For Euclid is indeed not the ultimate or last word on rigor, and there are notoriously what from our modern standpoint must be regarded as lapses rigor in the *Elements* (as seems to be conceded, somewhat grudgingly perhaps, by Adams in the passage quoted by Chihara), many of which were being repaired by contemporaries of Frege.

Hilbert's standard is that each theorem must follow logically from postulates stated in advance, where this means that logical form alone, regardless of subject matter or content, guarantees that *if* the postulates are true, *then* so is the theorem. Such is the meaning of the much-quoted statement attributed to Hilbert by his oldest graduate student in a biographical sketch at the end of the third and last volume of Hilbert's collected papers: One must at all times in place of "points, lines, planes" be able to say "tables, chairs, beermugs" (*man muß jederzeit an Stelle „Punkte, Gerade, Ebene" „Tische, Stühle, Bierseidel" sagen können*, Blumenthal 1935, p.403). Despite the great heuristic and pedagogical value of inspection of geometrical diagrams and deployment of visual intuition, Hilbert's standards rule out any appeal to such things *in proofs*, and not just because pictures

can sometimes be misleading (as in the "proofs" that every triangle is isosceles and some obtuse angles are right in Collingwood, 1899, pp. 900-902).

Euclid's practice does not always conform to these restrictions of Hilbert's. Most notoriously, Euclid lists the primitive constructions and assertions he is supposing to be allowed in Postulates 1-5, but then it seems immediately assumes another without explicit statement or acknowledgment in Proposition 1. This, it will be recalled, is the problem, given a segment *AB*, to construct an equilateral triangle *ABC* having it as a side. Euclid considers two circles with radius *AB*, the one with center at *A* and the one with center at *B*, whose construction is provided for by Postulate 3. But in subsequent reference to these circles he speaks of each not as the circle with such-and-such center and radius, but rather by mentioning three points on it. Now for any three non-collinear points there is indeed a unique circle passing through them, but Euclid does not deal with such matters until Book IV, so he may seem to be getting ahead of himself in Book I. Beere and Morison allude to this sort of thing as an illustration of the gap between existence and construction: by definition, every circle *has* a center, but to *find* the center when the circle is presented by naming three points is another matter. Crucially, by labeling his two circles "the circle *BCD*" and "the circle *ACE*" Euclid insinuates, without any logical justification on the basis of his announced postulates, that the two circles have a point *C* in common. In modern treatments this can be proved using a circle axiom to be stated shortly, but Euclid acknowledges no need for any such thing among his postulates.[2]

Besides the difficulty under discussion, which occurs not only in Proposition 1 but well beyond it, there are similar lapses involving the intersection of a line and a triangle, which by modern standards necessitate another missing axiom, introduced by Moritz Pasch in 1882. As the remark of Bernays hints, it was in the process of rigorization, filling in such gaps in Euclid's arguments as was done by Pasch and other nineteenth century figures culminating in Hilbert, that the explaining away of modal locutions in terms of existence statements occurred.

Euclidean straightedge and compass plane geometry was given its final rigorous form only in the twentieth century, by Alfred Tarski and coworkers, in a formal theory here to be called $\mathbf{G}_0$. (It is called $\mathrm{CG}^{(2)}$ in the authoritative survey of Tarski and Givant, 1999, p. 191.) The austere formalism of $\mathbf{G}_0$ is to begin with a *first-order* theory, one whose logical notions comprise those of standard classical logic as in present-day textbooks *and those only*, without any "higher-order" apparatus as in Hilbert, let alone intuitionist logical operators as in Brouwer and Heyting, or modalized quantifiers as in Chihara or Adams.

Moreover, the variables of the language are to be thought of as ranging only over points of the plane: there are no variables for lines or circles, for instance. And there are only two primitive predicates or relations symbols for two geometric relations:

(5a) $\mathbf{B}xyz$ or *betweenness*: x, y, z lie in a line, with y between x and z

(5b) $\mathbf{C}wxyz$ or *congruence*: w lies as far from x as y lies from z

(Here betweenness is to be understood inclusively, so $\mathbf{B}xxz$ and $\mathbf{B}xzz$ always hold, and identity of points $x = y$ is definable as $\mathbf{B}xyx$.)

Nonetheless, because any pair of distinct points may be regarded as *coding* a line (the one passing through both points) and a circle (the one having as center the first point and passing through the second point), one can express indirectly through coding the basic geometry of lines, circles, as well as many kinds of composite

[2] Is Euclid's practice in Proposition 1 a lapse not merely from the standards of Hilbert, but also a lapse from the standards of Euclid himself (of which we unfortunately have no quotable capsule formulation comparable to Hilbert on beermugs)? The existence of the needed point of intersection seems evident from the diagram if not from the text of Proposition 1, but leaving something that cannot be deduced from the text of a proof to be inferred from an accompanying diagram is not permitted by Hilbertian standards. What about Euclidean standards? For present purposes we may set aside this question, and more generally the whole contentious issue of the role of diagrams in ancient Greek mathematics, where they certainly play a much larger role than in Hilbert's work (with as an extreme case Euclid seemingly feeling obliged to attach a diagram to each proposition even in his arithmetical books, where they appear hardly more than doodles in the margin). A great deal discussion of considerable interest (see in particular Netz, 1998) has been written about Euclidean diagrams, but this discussion is not of direct relevance to the concerns of the present note. What we must acknowledge here is that *if* "Euclidean standards of rigor" do permit the sort of steps Euclid takes in Proposition 1, *then* Hilbert's standards do not only, as Frege's remark on Euclidean standards suggests, "go beyond them," but go *a very substantial distance indeed* beyond them.

figures. (Propositions about areas, including the Pythagorean theorem, admittedly present further challenges.) In particular, $\mathbf{G}_0$ has a *circle axiom* to the effect that if a line has both a point inside a given circle (of distance from the center less than the radius) and a point outside that circle (of distance from the center greater than the radius), it will have a point *on* the circle.

(Ellipses, too, can be coded, not by pairs but by triples of points, the two foci plus any point on the circumference. But in $\mathbf{G}_0$ there is no ellipse axiom comparable to the circle axiom, and arguments assuming the existence of intersections of conic sections could not have been carried out on the basis of Euclid's postulates—not that Euclid wished to go into such matters.)

Returning to (1)-(3), because there is no modal apparatus in the language of $\mathbf{G}_0$, (2) cannot be expressed if the modal locutions are understood literally or primitively, in terms of a possibility operator $\Diamond$. And if (2) is explained in terms of the existence of a program, it still cannot be expressed—not directly, since the variables do not range over programs; nor yet indirectly, since programs cannot be coded by pairs of points, as can lines and circles, or by configurations of any other fixed finite number of points.

But two more positive remarks can be made. There is a slight modification or small adjustment $\mathbf{G}$ of $\mathbf{G}_0$ of which the following may be said:

(A) By adding to the language of $\mathbf{G}$ symbols for certain functions definable in $\mathbf{G}$, corresponding to the familiar operations of finding the intersections of lines and circles, one can obtain a notation for straightedge and compass construction programs permitting the systematic expression of assertions of type (3).

(B) It can then be shown that, for any purely existential assertion of type (1) provable in $\mathbf{G}$, there is a construction assertion of type (3) that implies the given existenstial assertion and is provable in $\mathbf{G}$ taken together with the definitions of the new symbols

In the next two sections I will first sketch the construction justifying (A), then outline the proof justifying (B).

These results should not be surprising. On the contrary, they are just what one might expect (in both the factive and the normative sense) of a formalism in modern language that aspires to embody so far as possible the spirit of Euclid's construction-oriented geometry. But the departures from the *ipsissima verba* of ancient formulations needed to secure a level of rigor up to modern standards are numerous enough and large enough that (A) and (B) are not obvious without proof, either. Their proof is, so to speak, a check on the fidelity of Tarski and his school to something like the Euclidean point of view.

Before proceeding further let me indicate the "slight modification or small adjustment" needed in the formal geometry under consideration. For $\mathbf{G}_0$ as thus far described conspicuously fails to satisfy the most basic kind of "constructivist" requirement that every proof of the existence of a configuration of points with a certain feature has to have at least implicit in it a specification of a particular example of such a configuration.

The most trivial existential theorem is simply that there exists a point that is self-identical, or more simply still, that there exists a point. But it is not possible to specify any particular point using only the betweenness and congruence relations. This can be seen by looking at the most familiar model of $\mathbf{G}_0$, the *Cartesian plane* as studied in middle school. The points of the plane are pairs (a,a') of real numbers, and betweenness and congruence are defined by the usual formulas. Thus if $x=(a,a')$, $y=(b,b')$, $z=(c,c')$, $u=(d,d')$, $v=(e,e')$, then

(6a) $\mathbf{B}xyz$ holds if and only if

$a \leq b \leq c$ or $c \leq b \leq a$ and

$(b-a)\cdot(c'-a') = (b'-a')\cdot(c-a)$

(6b) $\mathbf{C}xyuv$ holds if and only if

$(b-a)^2+(b'-a')^2 = (e-d)^2+(e'-d')^2$

In this plane any condition Φ satisfied by any one point x will equally be satisfied by any other point x', so that the condition cannot be used to pick out a distinguished point. This is because there is a *translation* carrying x' to x, and translation preserves relations of betweenness and congruence, in terms of which Φ would be stated.

If we add a constant $\boldsymbol{a}$ to denote some specific, privileged or destinguished point a, the "constructivist" condition will still fail to hold. For it will be a trivial theorem that there exists some point y other than a, while again it is not possible to specify any particular example. This is because, given any point y distinct from a and any other such point y' there will be a rotation around a, keeping a fixed, that will carry y', if not yet to y itself, then at least to some other point y'' lying on the line ay on the same side of a as y. And then there will then be a *dilatation* (uniform expansion or contraction) leaving a fixed and carrying y'' to y. And rotation and dilatation, just like translation, preserve betweenness and congruence relations.

If we add a constant $\boldsymbol{b}$ to denote some specific, privileged or distinguished point b other than a, then there will be infinitely many points of the plane that *can* be uniquely specified in terms of the two distinguished points a and b. One of these will be the midpoint of the segment ab, characterizable as being between a and b and at the same distance from either. But all the specifiable points will, like this example, be on the line ab. And it is, of course, a theorem of plane geometry that the plane is more than one-dimensional, and so contains a point z not collinear with a and b. But for any such point z there is another such point z', though only one, satisfying all the same conditions involving betweenness, congruence, and the two distinguished point a and b, namely, the mirror image of z on the opposite side of ab. For there will be a transformation, namely, *reflection* in the line ab, leaving a and b fixed, that will carry z' to z, and reflections as much as translations, rotations, and dilatations preserve betweenness and congruence.

But if we add a constant c to denote some specific point c not on the line ab, then it is a theorem of plane geometry that the plane is less than three-dimensional, and there will be no two points z and z' satisfying all the same conditions involving betweeness and congruence and the three distinguished points; in particular there will be no two points simultaneously equidistant from a and from b and from c. From any pair of points z, z' we may therefore always specify one as what may be called *proximal* and the other *distal* relative to the three distinguished points a, b, c:

Namely, if the two points are at different distances from first point a, then the one nearer to a should be designated the proximal one. If the two points are equidistant from a but of different distances from the second point b, then the one nearer to b should be designated the proximal one. If the two points are equidistant both from a and from b, then they cannot be equidistant also from the third point c, and the one nearer to c should be designated the proximal one.

In what follows it will be convenient to pick as the point c one of the two points of intersection of the circle with center a passing through b and the line through a perpendicular to the line ab, thus obtaining what in the terminology of Burgess (1982) would be called a system of *benchmarks* a, b, c. In connection with such as system the point a may suggestively be termed the *origin*, and the lines ab and ac the *horizontal* and *vertical axes*, and b and c the *unit points* on those axes.

Now it is a completely general fact about first order theories that if we start with such a theory **T** in a language **L** in which can be proved $\exists x \Phi(x)$ for some condition expressed by a formula Φ of **L**, and if we then add a new constant $\boldsymbol{d}$ to **L** and the new axiom $\Phi(\boldsymbol{d})$ to **T**, the result **T'** will be what is called a *conservative extension* of **T**, meaning that anything statable in **L** (without the new constant) and provable in **T'** (with the new axiom) will be already provable in **T** (without the new axiom). (This is an immediate consequence of the rule of existential elimination found in textbook natural-deduction formulations of first order logic.) Further, by first-order logic, when any conclusion $\Psi(\boldsymbol{d})$ is provable in **T'** the following will be a theorem of **T**:

(7) $\quad \forall x(\Phi(x) \rightarrow \Psi(x))$

Similar results hold when adding two, three, or more constants. In particular, if we let **G** be the extension of $\mathbf{G}_0$ obtained by adding the constants $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and the axiom that the points they denote form a system of benchmarks, then the extension will be conservative, and to that extent harmless. (For it is a theorem of $\mathbf{G}_0$

there do exist systems of benchmarks, and indeed that for any two distinct points there exists a system of benchmarks of which the given points are the first two. Indeed, there exist exactly two such systems.) This technical adjustment will remove the three rather trivial failures of "constructivism" mentioned above. (A) and (B) should henceforth be understood as applying to this conservative extension **G** of $\mathbf{G}_0$, and the terms *proximal* and *distal* henceforth understood relative to the benchmarks denoted by the three new constants.

4. To Express a Construction Theorem

Where we have only points, as in Tarskian formulations, a straightedge and compass construction must be viewed as involving marking new points given previously marked points. (Constructing a line or circle will be a matter of constructing a pair of points coding one.) There are three basic kinds of steps:

(8a) to mark the intersection of the lines coded by x, y and by u, v

(8b) to mark the intersections of the line and circle coded by x, y and u, v

(8c) to mark the intersections of the circles coded by x, y and by u, v

Let now **L** be a first-order language and **T** a theory in **L** and Θ a formula of **L** for which it is a theorem of **T** that for any xs there exists a unique y such that Θ holds of the xs and y, or in symbols, the following:

(9) $\forall x_1 \forall x_2 \ldots \forall x_n \exists! y \Theta(x_1, x_2, \ldots, x_n, y)$

Then one can add an operator or function symbol $\boldsymbol{\vartheta}$ representing the function that, given any xs as input, gives their y as output. In this situation Θ is called the *defining formula* of $\boldsymbol{\vartheta}$, the relation expressed by $\Theta(x_1, x_2, \ldots, x_n, y)$ is called the *graph relation* of $\boldsymbol{\vartheta}$, while the defining axiom for $\boldsymbol{\vartheta}$, to be added to **T** when $\boldsymbol{\vartheta}$ is added to **L**, is the following:

(10) $\forall x_1 \forall x_2 \ldots \forall x_n \Theta(x_1, x_2, \ldots, x_n, \boldsymbol{\vartheta}(x_1, x_2, \ldots, x_n))$

Because the symbol $\boldsymbol{\vartheta}$ is definable, every formula containing it will have a formula in the original language **L** without the new function symbol that can be proved equivalent to it given **T** plus the defining axiom.

All this is true quite generally, for arithmetic, for geometry, for anything. In the specific context of geometry, we will want to add in this way to the language **L** of **G** function symbols corresponding to the basic construction steps (8abc). For this we must first indicate or recall how various auxiliary notions are expressible in the language of **G**. Here are a few basic ones:

(11a) $|xyz$ or *collinearity*:

"x, y, z lie on a line" or

"y lies on the line coded by x, z":

$\mathbf{B}zxy \vee \mathbf{B}xzy \vee \mathbf{B}xyz$

(11b) $\equiv xyz$ or *equidistance*:

"y lie at the same distance that z does from x" or

"y lies on the circle coded by x,z":

$x \neq y$ & $\mathbf{C}xyxz$

(11c) $\therefore xyz$ or *equilaterality*:

"x, y, z are the vertices of an equilateral triangle":

$\neg|xyz$ & $\mathbf{C}xyyz$ & $\mathbf{C}xzyz$

(11d) $\perp xyz$ or *perpendicularity*:

"the angle yxz is right":

$\neg|xyz$ & $\exists w(\mathbf{B}wxy$ & $\mathbf{C}xwxy$ & $\mathbf{C}zwzy)$ or equivalently

$\neg|xyz$ & $\forall w(Bwxy$ & $\mathbf{C}xwxy \rightarrow \mathbf{C}zwzy)$

(11e) $\oplus xyz$ or *benchmarking*:

"x, y, z form a system of benchmarks"

$\mathbf{C}xyxz$ & $\perp xyz$

(11f) $< xyz$ or *nearness*:

"y lies less far than z does from x":

$\exists w(w \neq z$ & $\mathbf{B}zwx$ & $\mathbf{C}xyxw)$ or equivalently

$\forall w(\mathbf{B}wzx \rightarrow \neg\mathbf{C}xyxw)$

(11g) $\langle tuvxy$ or *proximality*:

"t, u, v constitute a system of benchmarks and

of the two points x and y, the former is proximal

and the latter distal relative to them":

$\oplus tuv$ & $[< xyt$ or $(\mathbf{C}xtyt$ & $< xyu)$ or $(\mathbf{C}xtyt$ & $\mathbf{C}xuyx$ & $< xyv)]$

In **G**, where have distinguished a system of benchmarks denoted $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, we may write simply $\langle xy$ for $\langle \boldsymbol{abc}xy$. The list (11) could be indefinitely extended.

We next define five seven-place functions, one of them, $\boldsymbol{\alpha}$, related to constructions steps of kind (8a), two of them, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}$', related to construction steps of kind (8b), and two of them, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}$', related to construction steps of kind (8c). The benchmarks will play three roles in what follows.

First, constructions must begin with some data, as Euclid's Proposition 1 starts with there being given a line segment, or equivalently the two distinct points, its endpoints. The benchmarks will serve the starting points for all the constructions with which we will be concerned.

Second, before we can introduce a symbol $\boldsymbol{\vartheta}$for a function in the manner discussed in connection with (9) and (10), the function must be *total*, or defined for all inputs, even in waste cases where we do not care what the output is; and we may conventionally take the first benchmark as the waste-case output. For geometric constructions the definition of waste case for a function

$$\boldsymbol{\delta}(t,u,v,w,x,y,z)$$

where $\boldsymbol{\delta}$ is any of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$', $\boldsymbol{\gamma}$' will be a disjunction four clauses, the first three the same for all of (8abc):

(12a) t,u,v are not a system of benchmarks

(12b) $w = x$, meaning that the pair w,x fails to code a line or circle

(12c) $y = z$, meaning that the pair y,z fails to code a line or circle

The last clause will be different for each of the three basic construction steps, thus:

(13a) the lines coded by x,y and u,v do not intersect in a point

(they coincide or are parallel)

(13b) the line coded by x, y does not intersect the circle coded by u, v in two points

(the former is disjoint from or tangent to the latter)

(13c) the circles coded by x, y and u, v do not intersect in two points

(they coincide or are disjoint or are tangent)

In any waste case we will set the value of function $\boldsymbol{\delta}$ to be t.

Third, when we need to mark a point of intersection of a line or another circle with a given circle, we have two equally good options that are mirror images of each other, so to speak. Euclid's instructions for producing an equilateral triangle on a given base, as we find them implicit in his Proposition 1, do not tell us how to choose, and this may be said for many others of his construction problems as well: Buridan's ass would be unable to complete the construction given only Euclid's instructions. With benchmarks in the background we can agree that when we are faced with a choice between two points, we should always take the proximal one of that pair.

We can now write down the definitions of symbols for the functions connected with constructions (writing as usual in mathematics "iff" for "if and only if").

(14a) $\boldsymbol{\alpha}(t,u,v,w,x,y,z) = s$ iff we are in a waste case and $s = t$ or s is the point of intersection of the lines coded by w,x and by y,z

(14b) $\boldsymbol{\beta}(t,u,v,w,x,y,z) = s$ iff we are in a waste case and $s = t$ or s is the proximal point of intersection of the line and circle coded by w,x and by y,z

(14c) $\boldsymbol{\gamma}(t,u,v,w,x,y,z) = s$ iff we are in a waste case and $s = t$ or s is the proximal point of intersection of the circles coded by w,x and by y,z

(14d) $\boldsymbol{\beta}$' like $\boldsymbol{\beta}$ but for distal

(14e) $\boldsymbol{\gamma}$' like $\boldsymbol{\gamma}$ but for distal

Here in (14bcde) *proximal* and *distal* are to be understood relative to the benchmarks t,u,v.

Three trivial three-place functions as follows will also be wanted:

(15a) $\mathbf{i}(t,u,v) = t$ (15b) $\mathbf{j}(t,u,v) = u$ (15c) $\mathbf{k}(t,u,v) = v$

Applied to a system of benchmarks these pick out, respectively, the origin and horizontal and vertical unit points.

One further abbreviation will make for conciseness: Let $\S f g_1 \ldots g_n$ denote the m-place function obtained by substituting n given m-place functions in the n-place function f, thus:

(16) $\S f g_1 \ldots g_n(x_1,\ldots,x_m) = f(g_1(x_1,\ldots,x_m),\ldots,g_n(x_1,\ldots,x_m))$

Then a program for constructing a configuration of k points (starting from given benchmarks) can be represented as a k-tuple of complexes of the above-mentioned nine symbols $\boldsymbol{\alpha}$,$\boldsymbol{\beta}$,$\boldsymbol{\gamma}$,$\boldsymbol{\beta}$',$\boldsymbol{\gamma}$',$\mathbf{i}$,$\mathbf{j}$,$\mathbf{k}$ and §.

To apply this apparatus to Euclid's Book I, Proposition 1, we will want a complex ψ provably satisfying the following:

(17) $\therefore (\mathbf{i}(a,b,c),\mathbf{j}(a,b,c),\psi(a,b,c))$

Here **i** and **j** applied to our system of benchmarks will simply give us the first two of them, thus determining a line segment such as is given at the outset in Euclid's construction. The crucial ψ applied to our benchmarks should yield the proximal point of intersection of two circles, the radius of each being the distance between the first two benchmarks, and the centers of the two being the first and the second benchmark. This amounts to $\boldsymbol{\gamma}$ (**a, b, c, a, b, b, a**), which amounts to the composition of the seven-place function $\boldsymbol{\gamma}$ with the seven three-place functions **i, j, k, i, j, j, i**.

Hence the program for the construction of the vertices of an equilateral triangle, given benchmarks a,b,c can be represented by a trio of symbol complexes φ,χ,ψ as follows:

(18) $\varphi = \mathbf{i} \qquad \chi = \mathbf{j} \qquad \psi = \S\boldsymbol{\gamma abcabba}$

And the existence theorem that there is an equilateral triangle, namely,

(19) $\exists x \exists y \exists z \therefore (x,y,z)$

is implied by the following restatement of (17):

(20) $\therefore (\varphi(a,b,c),\chi(a,b,c),\psi(a,b,c))$

Leaving the benchmarks to be understood we may write this as $\therefore (\varphi,\chi,\psi)$ for short. And what has just been said about the equilateral triangle applies equally to the regular heptadecagon, substituting Gauss's construction for Euclid's. Behind a formalized version of (1) in the style of (19) there stands a formalized version of (3) in the style of (20).

More generally, for any condition expressible in a formula Φ, a statement of type (3), to the effect that an explicitly specified program would produce a configuration of points $x_1,x_2,x_3,\ldots$ satisfying Φ, can be taken to be a formula of the form

(21) $\Phi(\varphi_1,\varphi_2,\varphi_3,\ldots)$

wherein $\varphi_1,\varphi_2,\varphi_3,\ldots$ are symbol complexes of our notation for representing programs, indicating how points $x_1,x_2,x_3,\ldots$ are to be constructed from the benchmarks, applying specified cases of (8abc) in a specified order.

In sum, while a *constructibility* assertion of type (2), to the effect that there exists a program that would produce a given kind of configuration, has to be expressed in the "metalanguage," as the statement that there exist φs for which (21) is a theorem, by contrast we have produced (or at least sketched, relying on results of Tarski and his school) a way, namely (21) itself, of expressing in the "object language" a *construction* assertion of type (3), QEF.

5. There Is a Construction Theorem for Any Pure Existence Theorem

A *purely existential* formula (respectively, *purely universal* formula) is one that, when it is written out in primitive notation, with all defined symbols replaced by their definitions, consists of a string of existential quantifiers $\exists$ (respectively, universal quantifiers $\forall$) in front of a quantifier-free formula. (A bit confusingly, formulations of a theory that have only purely universal formulas as axioms are conventionally called *quantifier-free* formulations. This is because of the custom of omitting the initial universal quantifiers when a purely universal axiom is stated, leaving them to be tacitly understood. Thus for instance in arithmetic the commutative law for addition is expressed as

$$x+y=y+x$$

where what is really meant is the *universal closure* of this formula, namely

$$\forall x \forall y (x + y = y + x).$$

In effect, assertion of the version without the initial string of universal quantifiers is taken to be tantamount to assertion of the version with them.)

A formula provably equivalent in **G** to a formula that is purely existential (respectively, purely universal) is said to be of class Σ (respectively, class Π). A formula of class Δ is any that is of both classes Σ and Π, thus expressing a notion that can be given both a purely existential and a purely universal definition. The function denoted by a symbol $\boldsymbol{\vartheta}$ introduced as above is of class Δ if its defining formula is.

The basic features of these notions belong to general definability theory and include the following closure properties:

(22a) Σ is closed under $\wedge$ and $\vee$ and $\exists$

(22b) Π is closed under $\wedge$ and $\vee$ and $\forall$

(22c) Δ is closed under $\wedge$ and $\vee$ and $\neg$

and all three are closed under substitution of Δ functions. There is nothing specifically geometrical about these facts, nor about the fact that the negations of Σ formulas are Π formulas, and vice versa. Notably, the general proofs about Σ and Δ are very much like those for semi-recursive and recursive in computability theory (compare Boolos et al., 2007, p. 76, Theorem 7.4).

Now obviously (21) implies the existence statement

(23) $\exists x_1, \exists x_2, \exists x_3, \ldots \Phi(x_1, x_2, x_3, \ldots)$

which will be *purely* existential provided Φ is quantifier-free (and will be Σ provided Φ is Σ). This covers the case of the formula stating that the xs are the vertices of a regular polygon of n sides, for any $n \geq 3$. It also covers many other cases, for we will soon see that the class Σ and indeed the class Δ is quite large.

The main result whose proof will be outlined here is that for any purely existential theorem of **G** of type (23) there is a theorem of **G** of type (21) that implies it. (And note that we have already in effect seen in connection with (7) that if some statement about the distinguished system of benchmarks denoted $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ is a theorem of **G**, the corresponding generalization about all systems of benchmarks will be a theorem of $\mathbf{G}_0$.)

It is crucial for the proof of this result that the defining formulas Θ for our $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ can be taken to be Δ. This is verified by going through the list of defined notions in the preceding section one by one, and that list could be indefinitely extended. At the beginning, collinearity and equidistance and equilaterality were given quantifier-free definitions (11abc). Then perpendicularity and nearness were each given a Σ and an equivalent Π definition in (11df). The first problematic case comes only with (13a), or the parallelism of the lines coded by x, y and by u, v. Here the obvious definition is Π (the *non*-existence of a common point). The parallel postulate implies a less obvious equivalent definition that is Σ, in terms of the existence of points on the one line and on the other with the segment between them perpendicular to both lines. The other cases are left as exercises to the reader. The end result is that in a case of the kind that interests us, where Φ is Σ, the formulas of type (23) will be Σ, by the closure of that class under substitution of Δ functions and the other closure properties (22abc) of Σ and Δ.

The rest of the proof draws on the beautiful nineteenth-century algebra used to show that the classic problems of giving straightedge and compass constructions for duplicating the cube and trisecting the angle or constructing a regular heptagon are unsolvable (as expounded, for instance, in Papantonopoulou, 2002, chapters 11 and 12; the impossibility of squaring the circle, involving as it does the number pi, requires analytic methods). The *constructible field* is the smallest set of real numbers containing 0 and 1 and closed under the *rational operations* of addition, subtraction, multiplication, and division by non-zero numbers, as well as under extraction of square roots of positive numbers.

Tarski shows that a minimal model $\mathbf{M}_0$ of $\mathbf{G}_0$ can be obtained by taking as "points" pairs constructible numbers and defining the primitive relations just as in the Cartesian model, which is to say, just as in middle school analytic geometry. Square roots are needed to get the circle axiom. This $\mathbf{M}_0$ may be called the *constructible plane* and contrasted with the Cartesian plane in which **B** and **C** are defined by the same formulas (6ab), but the points are pairs of *arbitrary* real numbers. The points $a = (0,0)$, $b = (1,0)$, $c = (0,1)$ may be taken as benchmarks to get a minimal model **M** of **G**, and a crucial property of **M** is that *every point is straightedge and compass constructible from those benchmarks*.

This result has a long history. Euclid has in Book V a theory (which historians associate with Eudoxus) of ratios and proportions of lengths and other magnitudes, but he does not speak of these ratios as *real numbers* the way we would. Numbers for Euclid are positive integers. That is why it is an anachronism to speak of the Greek discovery of the incommensurability of the side and diagonal of a square as the discovery that $\sqrt{2}$ is an irrational number. In Euclid ratios are not things that can be added and multiplied.

But ratios *were* considered numbers at least as early as Omar Khayyam, *and constructions in Euclid could then be reconstrued as rational operations on ratios, as well as extraction of square roots*. This picture is in the background in Descartes' *Géométrie* and the foreground in Newton's *Universal Arithmetick*, and the theory is treated fully rigorously in Hilbert. It is also behind the results in Burgess (1984) on the possibility of reconstructing or reconstruing analytically-formulated theories of classical physics in a "synthetic" style.

The treatment of the straightedge and compass constructions related to multiplication, division, and square roots has found its way into the more thorough introductory textbooks of abstract algebra, as part of Galois theory. (See Papantonopoulou, 2022, Propositions 11.1.12 through 11.1.14, pp. 345-346.) In the constructible plane, any point $(x,0)$ on the horizontal axis can be obtained from $(0,0)$ and $(1,0)$ by the geometric steps corresponding to the algebraic steps used to obtain x from 0 and 1. Likewise any point $(0,y)$ on the vertical axis is constructible from the benchmarks $(0,0)$ and $(1,0)$. And then any point at all (x,y) is constructible from the benchmarks as the intersection of the horizontal line through $(0,y)$ with the vertical line through $(x,0)$.

Now suppose (23) is a theorem of **G**, with Φ quantifier-free or even just Σ. Then it must be true in the constructible plane, where every point is constructible, and if we take symbol complexes, φs, describing the construction of the pertinent xs, then (21) will be true in the constructible plane. But part of what was meant by calling **M** a "minimal" model is that *any* model of **G** has the constructible plane (or an isomorph or copy thereof) as a submodel. And it is a quite general fact of model theory, not at all tied to geometry specifically, that any Σ statement true in a submodel of a model is true in the model itself. Because (21) is Σ, it will thus be true in every model of **G**. And finally, by the Gödel completeness theorem it follows that it will be a theorem of **G**. And so we have shown (at least in outline) that behind every pure existence theorem there is a construction theorem, QED.

6. Related Issues

Let me take note now of a few questions about potential extensions or refinements of the results discussed so far.

First question. A proof has been outlined for the following "metatheorem": For every pure existence theorem *there exists* a construction theorem implying it, or more long-windedly, for every proof π of a pure existence theorem Φ, *there exists* a pair of proofs (ρ,σ) with ρ being a proof of a construction theorem Ψ, and σ being a proof of the logical implication $\Phi \rightarrow \Psi$. The question now arises: is the "metaproof" that has been given or sketched for this result "constructive" in the sense of that there is an effective procedure that applied to any given pertinent π will compute a suitable (ρ,σ)?

Yes, there is, and it goes like this: Nineteenth-century set theory established that given a finite alphabet of symbols, one can effectively list all finite strings τ of symbols from this alphabet in a sequence $\tau_0, \tau_1, \tau_2, \ldots$ indexed by natural numbers. Given a pertinent π, for any τ on the list one can effectively decide or compute whether it is a suitable pair (ρ,σ) for π. So just go down the list until a suitable pair is found. This procedure is "effective" or "computable" in the sense that in any given case it is *possible in principle* (not worrying about the amount of time needed to carry it out or the amount of space needed to record intermediate results), for a digital machine to carry it out.

But it is not *feasible in practice*. The function taking one from π to (ρ, σ) in the manner described is computable or recursive but might take astronomically long to compute. And this raises the question: *Is there any more efficient way to find a construction theorem behind any given existence theorem?* To begin with, could there be a procedure producing not just a recursive but a *primitive* recursive function? Such questions of "complexity theory" virtually always arise when something is proved to be computable or recursive. They are often quite difficult to answer, or anyhow, often go unanswered for a long time after they are originally posed. We need not be especially troubled, therefore, if the question just enunciated turns out to be one that has to be left open for the present and foreseeable future.

Second question. We may ask *how far can results similar or analogous to those established in this note for straightedge and compass construction be obtained for other kinds of geometric construction?* This is obviously not a single question but a series of them for different species of constructibility. Probably the most interesting case, and the only one I will explicitly discuss, is that of so-called *neusis* or *verging* constructions (known to be equivalent to *origami* constructions). Where straightedge and compass permit the duplication of the square and the bisection of an angle and the construction of a regular pentagon, they do not allow for the duplication of the cube or the trisection of an angle or the construction of a regular heptagon. These three things do have neusis or verging constructions.

By definition these are constructions using in addition to the compass a "marked ruler" or "notched straightedge." They are equivalent to constructions involving drawing of conic sections (using a string as mentioned in a note early on here in the case of the ellipse, or in some other way) and finding the intersections of such curves. They are equivalent also to constructions using a curve called *the conchoid of Nicomedes*, for the drawing or producing of which there is a mechanism whose invention is attributed by a commentator (one Eutocius of Ascalon) to Archimedes.

This class of constructions has been analyzed algebraically as thoroughly as have been straightedge and compass constructions. (See Papantonopoulou, 2002, pp. 357ff.) The algebraic counterpart of going beyond straightedge and compass constructions to neusis or verging constructions is in jargon that of going beyond "Euclidean" to "Vietan" fields, or in plainer language, from solving quadratic to solving cubic equations. This last is a problem that itself has long history, in which the outstanding figures are Omar Khayyam in the eleventh or twelfth century, for geometric solutions by intersecting conics, and Gerolama Cardano in the sixteenth century, for algebraic solutions by radicals.

There is much more that could be said, but what is most relevant here is that since the algebra has been so thoroughly analyzed it ought not to be very difficult to extend the methods of the present note to establish analogues of (A) and (B) for the wider class of constructions. Still, since I have not worked through the details, I can for the moment only advance this as a very plausible conjecture rather than claim it as a theorem.

Third question. How widely beyond the case of purely existential theorems can the notions and results of the present note be extended? What has been shown so far is that there is a construction theorem behind any purely existential theorem, that is, any theorem of form (23) above with Φ quantifier-free; and a moment's thought shows that this results extends to the case where Φ consists of a string of existential quantifiers follow by something quantifier-free. The next cases to consider would be those of *universal-existential* theorems, of the type

$$(24) \quad \forall x_1, \forall x_2 \ldots \forall x_m \ldots \exists y_1, \exists y_2 \ldots \exists y_n \Phi(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$$

with Φ quantifier-free.

The methods used here do establish *some* results of this kind. For instance, it can be shown not just how to construct an equilateral triangle but how for any line segment to construct an equilateral triangle *having that segment as a side* (which is the actual result in Euclid's Proposition 1). For the construction given above establishes that there is an equilateral triangle having the segment *ab* joining the first two benchmarks as one side, and we have already seen that what is provable in **G** to hold for a, b, c, is provable in $\mathbf{G}_0$ to hold for *any* system of benchmarks, and that for any two distinct points there is a system of benchmarks of which they are the first and the second. The general case is, however, so far as I know open.

The next case to consider would be that of *existential-universal* theorem, of the type

(25) $\exists y_1, \exists y_2 \ldots \exists y_n \forall z_1, \forall z_2 \ldots \forall z_p \Phi(y_1, y_2, \ldots, y_n, z_1, z_2, \ldots, z_p)$

with Φ quantifier-free. But it can be shown that there is *not* always a construction theorem behind such a result. It will be worthwhile to spell out a counterexample, since it will illustrate an important point about the distinction between Euclid-type construction-oriented geometry and Brouwer-type constructivist or intuitionist logic, a theme that thus far has been left in the background here since it was initially enunciated. For the counterexample will depend on the law of excluded middle.

It is known that *proportionality* or equality of ratios of segments, written in traditional notation thus:

$$st : uv :: wx : yz$$

is expressible in the language of **G**, and indeed is of class Δ. Hence it is expressible that the quintuple $\boldsymbol{u}$ of points u_1, u_2, u_3, u_4, u_5 are so related that

(26a) they all lie on the same line and in the order listed

(26b) u_1 and u_5 lie at the same distance from u_2 on opposite sides, so that u_1u_2 and u_1u_5 are in the ratio 1 to 2

(26c) u_1u_2 is to u_1u_3 as u_1u_3 is to u_1u_4 and as u_1u_4 is to u_1u_5, so that u_1u_2 and u_1u_3 are in the ratio 1 to $\sqrt[3]{2}$

Since $\sqrt[3]{2}$ is the number that has to be constructed in the problem of the duplication of the cube, when (26abc) hold for u_1, u_2, u_3, u_4, u_5 let us say the quintuple $\boldsymbol{u}$ is a *cube-duplicator*, and write $\mathbf{Q}\boldsymbol{u}$. Similarly write $\mathbf{P}\boldsymbol{u}$ to express that the quintuple $\boldsymbol{u}$ is *pentagonal* in the sense that u_1, u_2, u_3, u_4, u_5 in that order are the successive vertices of a regular pentagon. Let us also write $\exists\boldsymbol{u}$ to abbreviate $\exists u_1 \exists u_2 \exists u_3 \exists u_4 \exists u_5$. Then the counterexample to be discussed will be statable in this abbreviated notation thus:

(27) $\exists\boldsymbol{u}(\mathbf{Q}\boldsymbol{u} \vee (\mathbf{P}\boldsymbol{u} \,\&\, \neg\exists\boldsymbol{v}\mathbf{Q}\boldsymbol{v}))$

This is logically equivalent to the following existential-universal formula:

$$\exists\boldsymbol{u}\forall\boldsymbol{v}(\mathbf{Q}\boldsymbol{u} \vee (\mathbf{P}\boldsymbol{u} \,\&\, \neg\mathbf{Q}\boldsymbol{v}))$$

It is provable in **G** that $\exists\boldsymbol{u}\mathbf{P}\boldsymbol{u}$. It is neither provable nor disprovable in **G** that $\exists\boldsymbol{u}\mathbf{Q}\boldsymbol{u}$, for there is a cube-duplicator in the full Cartesian plane but not in the restricted constructible plane. To prove (27), note that by excluded middle (in any given model) there either does or does not exist a cube-duplicator. If there does, any cube-duplicator may be taken for the $\boldsymbol{u}$ in (27) and the disjunction will hold because its first disjunct does. If there does not, then the second conjunct of the second disjunct in (27) will hold, and if we take for $\boldsymbol{u}$ any regular pentagon—and we know there will be one—the first conjunct will hold as well, and hence the second disjunct as a whole, and hence the disjunction as a whole. So either way, (27) holds. The argument here is only slightly fancier than that used for (4).

Now consider any quintuple $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, \psi_4, \psi_5)$ of operations compounded out our basic operations $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$, and so on, and apply it to the benchmarks to obtain the quintuple of points

(28) $\psi_1(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}), \psi_2(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}), \psi_3(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}), \psi_4(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}), \psi_5(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$

Part of what was meant by calling the constructible plane a "submodel" of the Cartesian plane is that when applied to the benchmarks or any other elements of the smaller plane, $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$, and the rest *will give the same results whether we are thinking of them as operators on that plane or on the larger plane*. So the notation (28), which may be abbreviated as or $\boldsymbol{\psi}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$, or even just as $\boldsymbol{\psi}$, is unambiguous. A construction theorem implying the existence theorem (27) would have to look like this:

(29) $\mathbf{Q}\boldsymbol{\psi} \vee (\mathbf{P}\boldsymbol{\psi} \,\&\, \neg\exists\boldsymbol{v}\mathbf{Q}\boldsymbol{v})$

But no such thing can be a theorem of **G**, since nothing like (29) can be true in both the constructible and the Cartesian plane. To be true in the smaller plane, where there are no cube-duplicators and hence the first disjunct is false, the second disjunct and in particular its first conjunct, must be true. But that means that ψ is giving us a regular pentagon. By contrast, to be true in the larger plane, where the second conjunct of the second disjunct, and hence the second disjunct as a whole, is false, the first disjunct must be true. But that means that ψ is giving us a cube-duplicator. And nothing is both a regular pentagon and a cube-duplicator.

So while the main results (A) and (B) above hold for *purely* existential theorems as was seen in §3, we have just seen this success does not extend to *arbitrary* existence theorems. One can only conclude that **G** is "constructive" in one sense and "non-constructive" in another sense. But that was true already of Euclid and does *not* in itself demonstrate infidelity of the Tarskian approach to the Euclidean.

7. Not Unrelated Developments

At the suggestion of a reader of an earlier draft of this note, let me before closing describe some previous work on constructivity in geometry by Nancy Moler and Patrick Suppes (1968). It is clearly relevant to the present discussion, even though its results do not directly quite yield (A) and (B) above, while inversely and more emphatically, what is done here certainly does not accomplish the aims of the joint authors' more ambitious project. A certain gap or space exists between the two investigations, which in some respects are even opposite in their perspectives.

Here, in attempting to modernize the Euclidean context of Beere's and Morison's historical discussion of existence theorems versus construction problems, I have simply adopted Tarski's axiomatization, whose primitives are predicates, and whose postulates include items (notably Pasch's axiom) of universal-existential form. The tradition originating with Moler and Suppes, by contrast, seeks to do for geometric theories what has already been done for any number of arithmetic or algebraic theories, namely, to formulate them with only function symbols rather than predicates as primitives—constants are also allowed, since they may be counted as zero-place function symbols—and with all axioms as so-called quantifier-free (really, purely universal) statements. It might not be too inaccurate to say that whereas the interest here has been in finding construction theorems "behind" existence theorems, the interest in the tradition under discussion is more in simply avoiding existence assertions in favor of construction assertions.

Moler and Suppes are aiming to give a different axiomatization of what is in some sense supposed to be the "same" geometric theory as the Tarskian formulation they contrast with it. In the end they establish the agreement of their proposal with earlier models by means of *representation theorems*, which play a large role elsewhere in Suppes' *œuvre*. It is, however, also possible to give a more syntactic definition of the sort of thing that is wanted:

(30a) The function symbols of the new theory are to be definable in terms of the predicates of the old theory.

(30b) The axioms of the new theory are to be deducible from those of the old theory together with the definitions just mentioned of the function symbols.

(30c) The old predicates are to be definable in term of the function symbols of the new theory.

(30d) The axioms of the old theory are to be deducible from those of the new theory together with the definitions just mentioned of the predicates.

What is meant in (30a) is definability in exactly the same sense in which α, β, γ here were defined in terms of **B** and **C** and a, b, c. It should be mentioned that it turns out to be needful, in order to avoid trivialization, for Moler and Suppes also to introduce constants a, b, c denoting three non-collinear points, like the benchmarks in this note. Once this is done there is a very general way, not specifically geometrical, and perhaps not optimal, but always available, to introduce function symbols $r, s, t, \ldots$ for the new theory that will be interdefinable with the predicates **R,S,T**,... of the old theory.

How this may be done is sufficiently illustrated by the case of a single two-place predicate **R**. We associate with the predicate a symbol **r** for (a version of) the *characteristic* or *indicator* function *r* for **R**, which for given

x and y as inputs returns as output the item denoted $\boldsymbol{a}$ if $\mathbf{R}(x,y)$ holds, and that denoted $\boldsymbol{b}$ if not. Replacing $\mathbf{R}(x,y)$ by $\boldsymbol{r}(x,y) = \boldsymbol{a}$ will be enough to enable one to accomplish (52cd), and inversely, accomplishing (56ab) can be carried out by the usual methods of eliminating function symbols (as in Boolos et al. 2007, §19.4, pp. 255-258).

hat usual method will involve introducing some existence axioms like (9). Moving in the opposite direction will make such axioms superfluous, but will not by itself eliminate all the *old* existence axioms already present in the original version of the theory. It is, however, as mentioned in passing above, the avowed aim of Moler and Suppes in their geometric context to do just that: eliminate all existential axioms from the new theory, finding an axiomatization using only universal formulas. As they say (p. 143)

> In view of the highly constructive character of Euclidean geometry, it seems natural to strive for a formulation that eliminates all dependence on purely existential axioms…

And there exists what may superficially appear to be an all-purpose method for eliminating existential axioms at the cost of introducing new function symbols, a method of which some readers will have heard, called "reduction to Skolem normal form" or simply "Skolemization" (as in Boolos et al. 2007, §19.2, pp. 247-253). But in fact Skolemization is a red herring, not really relevant to the Moser-Suppes project, or to that of this note. (And readers unfamiliar with it may simply skim or skip the next paragraph here, which explains why.)

In general, Skolemization does not even guarantee that the new functions introduced will be definable in terms of what was there before their introduction, as required by (30a). That aside, it is certainly not something Moler and Suppes wish to rely on, since they continue the passage just quoted with these words:

> …but not, of course, by use of some wholly logical, non-geometric method of quantifier elimiination.

And Skolemization is the very paradigm of a wholly logical, non-geometric method of quantifier elimination. Rather, Moler and Suppes call for primitive function symbols representing "familiar constructions," of the kind whose use is acknowledged in the tradition of geometric construction problems, and something like this requirement of connection with historically given functions or operations has also been implicitly or tacitly assumed in the approach taken in this note.

Moler and Suppes take as their "familiar" geometric constructions *finding the intersection of lines*, a version of which was the first (8a) of the basic constructions (8) used in §3 above, and *the laying out of segments* as in Euclid, Book I, Proposition 2. Here the aim has been similar, to use only what are recognizably versions of the familiar construction steps (8abc).[3] Moler and Suppes do not concern themselves with finding the intersection of circles as was done in this note with (8bc), because they are working not with the Euclidean geometry that has been the focus here, but with the weaker Pythagorean geometry, which lacks the circle axiom.

The difference between the two geometries parallels the difference in algebra between Euclidean ordered fields, in which every positive element has a square root, and Pythagorean fields, satisfying the weaker condition that every sum of squares has a square root. The Euclidean as opposed to Pythagorean case was taken up later, by Horst Seeland (1978). By the time we come to the survey of Victor Pambuccian (2008), multiple geometries have come into play (including the stronger geometry of neusis or verging constructions alluded to earlier, and the oldest kind of *non*-Euclidean geometry, the hyperbolic), and the bibliography of this tradition has grown

[3] That the primitives such as $\boldsymbol{\beta}$ used here are importantly different from the ordinary straightedge-and-compass construction of the intersections of a lines and circles is evident and must be acknowledged. Tibor Beke (personal communication) has in this connection pointed to the difficulty, which I take to be the impossibility, of determining by ordinary staightedge-and-compass constructions which of three collinear points lies between the other two. This *can* be determined if one is able to determine, given any two of the points, which is nearer the third. For if C and B are equidistant from A, then A is the in-between point; and otherwise, if say C is further away than is B from A, then C cannot be the in-between point, and hence A and B must lie on the *same* side of C, in which case whichever is nearer to C is the in-between point. Now by ordinary straightedge-and-compass constructions one can obtain the line on which the three points lie, and on that line the midpoint of the segment AB, and therewith obtain the circle with center on that line passing through both points A and B. With the distinction made by $\boldsymbol{\beta}$ between the *proximal* and *distal* points of intersection of the line in question with the circle in question, relative to a system of benchmarks that may be freely chosen, and in particular so chosen that C is the origin, the proximal point of intersection will be the nearer to and the distal point of intersection the farther from that origin, and the problem can be solved.

to nearly 150 items. There is a great deal here that might be looked over again from the somewhat different perspective of this note.

Moler and Suppes encounter the difficulty, which they describe as "akin to division by zero," that their functions are not always defined (as the intersection of lines does not exist when the lines are parallel). Indeed, they cite such difficulties as probably the main reason why something like their project was not carried out much earlier in the history of formalized geometry. Here such difficulties were dealt with by conventionally assigning a sort of null value, the origin, to waste cases where this happens. Moler and Suppes take the bolder line of simply allowing some functions in some models to be partial, and generalizing model-theoretic notions such as submodel and isomorphism to apply to this wider-than-usual range of models.

The paper of Moler and Suppes was published in a journal founded by Brouwer after Hilbert had, at a sort of climax of the *Grundlagenstreit*, manœuvred him off the editorial board of the *Mathematische Annalen*. And in a footnote at the bottom of the first page of their work they refer to Brouwer's chief disciple and tell us the following:

> It is a pleasure to dedicate this paper to Professor Heyting on the occasion of his seventieth birthday. In view of his long interest in constructive mathematics and in geometry, we believe the subject of our paper make it particularly appropriate to dedicate it to him.

For there is indeed a long tradition among intuitionists of concern with geometry, from parts of Brouwer's dissertation onward, through work by Heyting partly on problems to which Brouwer directed him, and then on to Heyting's student Dirk van Dalen and beyond. Still. overlap with the Moler-Suppes tradition may be somewhat limited insofar as the latter sticks to classical logic.

A genuine intuitionist would not allow definitions by cases, as repeatedly used here (in particular, in the definitions of α, β, γ) unless the case hypotheses are decidable. And just as in intuitionistic algebra the identity of real numbers is not assumed decidable, so also for the coincidence of planar points in intuitionistic geometry. There exists a large and more recent body of work produced by Michael Beeson, culminating in Beeson (to appear), specifically concerned with geometry that is "constructive" in the double sense of being *both* occupied with straightedge and compass constructions *and* based on Brouwer's and Heyting's intuitionistic or constructivistic logic. But this work inevitably has a rather different flavor from the work discussed or reported here.

References

Beere, J., & Morison, B. A mathematical form of knowing how: The nature of problems in Euclid's geometry. Unpublished manuscript.

Beeson, M. Constructive geometry. To appear. Available at http://www.michaelbeeson.com/research/papers/ConstructiveGeometryFinalPreprintVersion.pdf

Bernays, P. (1964). On platonism in mathematics. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics: Selected readings* (pp. 274-286). Englewood Cliffs, NJ: Prentice-Hall.

Blumenthal, O. (1935). Lebensgeschichte. In D. Hilbert, *Gesammelte Abhandlungen, Dritter Band* (pp. 388-429). Springer.

Boolos, G. S., Burgess, J. P., & Jeffrey, R. C. (2007). *Computability and Logic* (5th ed.). Cambridge University Press.

Burgess, J. P. (1984). Synthetic Mechanics. *Journal of Philosophical Logic*, 4, 379-395.

Chihara, C. (1973). *Ontology and the Vicious Circle Principle*. Ithaca, NY: Cornell University Press.

Collingwood, S. D. (Ed.). (1899). *The Lewis Carroll Picture Book*. London: Collins.

Frege, G. (1953). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number* (J. L. Austin, Trans.; 2nd rev. ed.). New York, NY: Harper & Brothers.

Heath, T. (1968). *The Thirteen bottomks of Euclid's Elements* (2nd ed.). Cambridge: Cambridge University Press.

Hilbert, D. (1902). *The Foundations of Geometry*. LaSalle: Open Court.

Moler, N., & Suppes, P. (1968). Quantifier-free axioms for constructive plane geometry. *Compositio Mathematica*, 20, 143-152.

Netz, R. (1998). Greek Mathematical Diagrams: Their Use and Their Meaning. *For the Learning of Mathematics*, 18, 33-39.

Netz, R. (1999). Proclus' division of the mathematical proposition into parts: How and why was it formulated? *Classical Quarterly*, 49, 282-303.
Pambuccian, V. (2008). Axiomatizing geometric constructions. *Journal of Applied Logic*, 6, 24-46.
Papantonopoulou, A. H. (2002). *Algebra: Pure and Applied*. Upper Saddle River: Prentice-Hall.
Seeland, H. (1978). *Algorithmische Theorien und konstruktive Geometrie*. Stuttgart: Hochschulverlag.
Tarski, A., & Givant, S. (1999). Tarski's system of geometry. *Bulletin of Symbolic Logic*, 2, 175-214.
Weisstein, E. Heptadecagon. Available at http://mathworld.wolfram.com/Heptadecagon.html