Euclid's Proof of the Infinitude of Primes: Distorted, Clarified, Made Obsolete, and Confirmed in Modern Mathematics

REINHARD SIEGMUND-SCHULTZE

In this article[1] I reflect on the recurrent theme of modernizing historical mathematical proofs, vocabulary, and symbolism, and the extent to which this modernization serves to clarify, is able to preserve, or is bound to distort the original meaning. My example is also a recurrent one: Euclid's proof of the infinitude of primes in book IX, theorem 20 of his *Elements*. Elementary number theory is a very appropriate field for discussing such general historiographic questions on a nontechnical level.[2] Indeed, a widely quoted and stimulating article by Michael Hardy and Catherine Woodgold (henceforth H&W) entitled "Prime Simplicity" appeared in this journal in 2009. H&W discuss Euclid's proof mainly from the point of view of its distortion by modern authors who quite often claim that it had the form of a *reductio ad absurdum*.[3]

My article, on the other hand, aims at a systematic presentation and logical analysis of Euclid's proof (as it is preserved in the critical editions by Euclid scholars) and, above all, at a more detailed discussion of its interpretation

---

[1] I dedicate this article to Walter Purkert (Bonn), the coordinator and very active force behind the excellent Felix Hausdorff edition soon to be completed, on the occasion of his 70th birthday in 2014.

[2] Meštrović ([2012]) contains many details on modern proofs of the infinitude of primes but is less interested in Euclid's original theorem and its modern variants. Most of his material goes far beyond the simple fact of the infinitude of primes that is the focus of the present paper.

[3] In the following, the notions *reductio ad absurdum*, proof by contradiction, and indirect proof are used interchangeably for assuming the logical opposite of what one wants to prove and then reaching, by legitimate logical steps, a contradiction that requires the assumption of the original claim.

and rewriting by a few select modern mathematicians, foremost among them Dirichlet, G. H. Hardy, and Hilbert.[4]

As I will show, Euclid's proof was constructive. For all the emphasis put on properties and axiomatic structure as displayed in Euclid's *Elements*, constructions—finding, in finitely many legitimate steps (i.e., those allowed by axioms and theorems), a mathematical object with certain properties—is the backbone of Greek mathematics.[5] More specifically, I will argue that Euclid's proof is "weakly constructive," and it is partly this weak form of constructiveness—the fact that Euclid does not provide an effective method (formula) to calculate new primes from given ones—which provided modern mathematicians with the wrong impression that Euclid's proof was indirect.

I agree with H&W that the insinuation of an indirect proof in Euclid IX, 20 became the "prevailing doctrine" when modern mathematicians wanted to relate in a positive way to Euclid's geometry. However, modern presentations of Euclid's proof, such as the ones by Hardy in 1938 and 1940, are *not* genuine indirect proofs of the infinitude of primes. They are quite close in spirit and content to Euclid's original proof.

What Does ''Projecting Modern Mathematics into the Past'' Mean, and How Much Historical Accuracy Can or Should One Reach in Modern Texts?

Mathematicians and historians projecting modern mathematics into the past can miss essential points of the intentions and results of the historical creators of mathematics. To mention a few: the modern notion of functional dependence and its graphic representation originated (basically) in the Scientific Revolution of the 17th century, connected to the need to extend the realm of mathematical objects to some "mechanical" curves (higher algebraic curves treated by Descartes) and to formulate the laws of physics (Newton). Projecting that modern notion uncritically into the past does not, however, help us much to understand Apollonius's sophisticated theory of conics. Even Euclid's *Elements*, which, because of their axiomatic structure, have a more familiar appeal to modern mathematicians than Apollonius's works, cause heated methodological-historical debates in their modern interpretations. For example, the deep theory of geometric proportions in book V of the *Elements* caused R. Lipschitz, in his discussion with R. Dedekind in the 1870s, to claim its logical equivalence with the modern definition of real numbers. The latter, being the inventor of the Dedekind cut, of course denied that claim, and most historians agree with him (Nikolić 1974). One consequence is that the claim

that the Ancients "proved the irrationality of the square root of two" cannot be unqualifiedly maintained either. More recently, discussions were raised about the so-called "geometric algebra," that is, the claim (for instance defended by B. L. van der Waerden and meanwhile refuted by historians) that some of Euclid's geometric theorems have to be interpreted as translations of historically preexisting algebraic problems and equations.[6]

Historians know, of course, that absolute historical "accuracy" cannot be attained in modern presentations and that they have to refer to modern vocabulary to make the very process of creation understandable.[7] In discussing Euclid's *Elements*, the problem starts with the question as to what extent it is legitimate to replace the purely verbal original formulation throughout the book by some other, modernized one that uses symbols. In the special case of the Pythagorean-Euclidean number theory, Euclid's dressing up of the theorems in geometric clothing, without genuinely geometric content, adds to the problem.

In the case of Euclid's *Elements*, one is best served relying on Bernard Vitrac's careful French edition (1990–2001) with very detailed commentary,[8] which is still based (as is the most famous English edition by Thomas Heath from 1908 and 1925) on the Greek text, edited by the Danish historian J. L. Heiberg in 1883–1885.[9] Vitrac says in his edition that—because of a scarcity of original sources—each modern specialist of Greek number theory provides a different interpretation (Vitrac 1994: 288). This should warn us, the nonspecialists, to restrict our discussion basically to the modern reception of what the specialists agree is a reliable text of Euclid's theorem IX, 20 and other theorems related to it.

Vitrac's and Heath's commentaries (and others) do not contribute many answers to my questions: to what extent is Euclid's theorem IX, 20 constructive or its proof indirect, what kind of infinity does it claim, and how has the theorem been interpreted by modern mathematicians. Nevertheless I will try to use their and other Euclid scholars' expertise to provide nuancing counterbalance against some formulations that necessarily have to be somewhat simplifying so as to be short and understandable.

My first question is admittedly speculative and cannot be answered in this article (although an answer would shed light on our problems): *Why did Euclid prove the "infinitude of primes" in the first place?*

The proof comes in book IX of the *Elements*, as proposition 20, almost at the end of the "number-theoretic" ("Pythagorean") books VII through IX, whereas the decisive instruments and propositions, particularly the Euclidean algorithm and propositions close (though not

---

fully equivalent) to the fundamental theorem of arithmetic, are presented in the beginning and in the middle of book VII. The algorithm and the propositions are frequently used in Euclid's number-theoretic books, but his theorem on the "infinitude" of primes is not used at all. The remaining propositions of book IX, 21 through 36, are mainly about the properties of odd and even numbers.[10] Thus theorem IX, 20 appears isolated. Did Euclid see it as the demonstration of a kind of ideal, platonic existence of infinitely many primes, something like the last theorem of the *Elements*, proposition XIII, 18, in which he proves that there exist no more than 5 platonic polyhedra?[11] That might explain why a "weakly constructive" proof sufficed for him: he apparently needed the additional primes only in a general "philosophical" sense, not to prove other theorems.

What Did Euclid Really Claim and Prove and What Is Its Modern Counterpart?

Euclid does not really claim or prove the infinitude of primes. What he says and proves in IX, 20 is:

ECL (Euclid's claim):

"Prime numbers are more than any assigned multitude of prime numbers."

It may appear pedantic and trivial to stress,[12] but it is historically a relevant fact that it requires a very simple, yet indirect argument to derive from this statement the following modern claim:

MCL (Modern claim of the infinitude of primes, in the Cantorian sense of actual infinity):

"There exist infinitely many primes."

To derive the MCL we make the (wrong) assumption that the finite "assigned multitude" is "complete" (contains all natural numbers that fall under the definition of a prime number). That is, the (finite) "assigned multitude" must contain all primes up to a maximum, which is *not* stipulated in Euclid's original proof. This explains why most modern presentations of Euclid's proof use uninterrupted sequences of primes (beginning with 2) and their products (primorials) as starting points. Although the infinitude follows easily from contradiction, one could also argue that this conclusion of infinitude can be drawn "constructively" in the sense of a continued counting of a *potentially infinite* set of prime numbers.[13]

Modern misinterpretations of ECL in the sense of MCL lead to the insinuation that Euclid's proof was basically or

substantially by contradiction. Various forms of **MIP (Modern Infinitude Proofs)** sometimes are substantially indirect proofs (as in the cases of Dirichlet [1863] and Hardy [1908] to be discussed later) and sometimes claim to be indirect although they follow closely the Euclidean original (as in the case of another proof by Hardy of 1938 and 1940).

To be sure, indirect proofs were not foreign to the Greeks (they rather invented them) and many Greek philosophers/scientists (among them mathematically highly educated individuals such as Plato and Eudoxus, who stood behind Euclid's *Elements* with their spirit and/or results) speculated about infinity.[14] André Weil, who was deeply interested in the history of number theory, pointed to book X, def. 3 of the *Elements* to show that Euclid was not averse in principle to talking about infinity. At the same time, Weil uses that example to caution readers not to expect too much impact of the philosophical positions of Greek mathematicians on their work:

> The views of Greek philosophers about the infinite may be of great interest as such; but are we really to believe that they had great influence on the work of Greek mathematicians? Because of them, we are told, Euclid had to refrain from saying that there are infinitely many primes, and had to express fact that differently. How is it then that, a few pages later, he stated that "there exist infinitely many lines" incommensurable with a given one?[15]

Thus one may never fully know whether Euclid added in his mind such an indirect proof, leading him from ECL (which he proved) to MCL. But this question does not affect at all the problem of whether his proof, as we have inherited it, was (logically) indirect, that is, by contradiction.

What we know for sure, however, is Euclid's proof of ECL and that it was done by a (weak) construction that can be considered as performing the following assignment:

ECO (Euclid's construction):

"Given a finite number of primes, find at least one additional prime."

Constructions occur in almost all theorems in Euclid's *Elements*, at least as auxiliary methods. Constructions are usually also applied in proofs of theorems that do not ask for constructions or that do not talk about the existence of mathematical objects, but only about properties. The

---

[10]This was also observed by David E. Joyce in his online version of Heath's English edition, which is quoted later. Vitrac (1994: 86), calls these last theorems ''enigmatic,'' meaning that their connection to the preceding theorems of the arithmetic books is problematic anyway.

[11]This comparison seems also relevant with respect to our central question of constructiveness: the construction of the five platonic polyhedra in propositions 13 through 17 of book XIII requires incomparably more effort than the ''negative'' result in proposition 18 that there are no more than the five regular polyhedra. Like Euclid's theorem on the infinitude of primes, his last theorem in book XIII, 18 obviously remains without application in the *Elements*, because it comes last. Vitrac (1994: 273) cautions against jumping too easily to the assumption that Euclid was a Platonist.

[12]We do not go quite so far though as putting ''proof of the infinitude of primes'' into quotation marks in the title.

[13]The resulting infinitude would, of course, not necessarily be complete. Cf. for instance Graham/Knuth/Patashnik (1989: 108) where such an infinitude is recursively constructed with the help of the Euclidean algorithm and the auxiliary notion of relatively prime ''Euclid numbers.'' Vitrac (1994: 445) stresses that this infinity is fully compatible with Aristotle's notion of potential infinity.

[14]However, in mathematics itself they usually treated infinity indirectly (as for instance in Eudoxus' famous notion of proportion in book V of the *Elements*) and preferred the safety of the finitely many steps in mathematical constructions. Of course this is related to the fact that the Greeks did not possess modern analysis.

[15]Weil (1978: 230). Knorr (1975: 233) refers to the remark in Euclid as an ''anomaly,'' because it is not a definition but seems to promise a theorem for which one has, however, to ''wait in vain.''

majority of theorems in the *Elements* are of this kind.[16] The most famous example is probably the "theorem of Pythagoras" in book I, theorem 47. Some propositions, however, are explicitly formulated as construction assignments, and the latter—combined with their solution that then follows—have to be understood as "construction theorems." The best known of this kind in the "number-theoretic" books VII–IX is the so-called "Euclidean algorithm," namely proposition VII, 2:

"Given two numbers not prime to one another, to find their greatest common measure."[17]

Expressed in modern words, in Euclid's number theory, constructions are usually performed by the four fundamental operations of arithmetic, in particular addition and multiplication, which in the *Elements* appear in geometric clothing.[18] But also in the number-theoretic books one finds some tendency to hide constructions in favor of properties. Thus ECO is not formulated separately in the *Elements* but it appears as the proof method in proposition IX, 20 with its claim ECL. I will argue in the following also that the use of the notion of the least common multiple in ECO instead of the more constructive notion of a product seems to indicate this predilection for logical structure and contributes to the underestimation of the constructive character of Euclid's theorem of the infinitude of primes.

[smaller] number") by "product" of numbers. Both notions occur in Euclid; they are differently defined (Vitrac 1994: 258), the first based on properties, the second on construction, but they are logically equivalent.

Euclid's proof is then based on two simple lemmas (L1 and L2), and on one not-too-deep theorem (T). Whereas theorem T is proposition VII, 31 in Euclid's *Elements*, the two lemmas do not explicitly appear in the *Elements*, but they follow easily from the notion of the product of natural numbers.

L1: The product of any two numbers (none of them the unit) is bigger than either of them.

L2: A product of n numbers (none of them the unit) + 1 is not divisible by any of these numbers.

T: Any product of numbers (none of them the unit) is divisible by a prime number.

L2 refers to a legitimately constructed number $P + 1$, which is the main idea of the proof, T continues the construction based on the possibility of repeated division.

Euclid, *Elements*, book IX, proposition 20:
ECL: "Prime numbers are more than any assigned multitude of prime numbers."
Proof (ECO):[19]

Let A, B, and C be the assigned prime numbers.

I say that there are more prime numbers than A, B, and C.

Euclid's Proof of ECL in IX, 20 Based on ECO

It is necessary to recall Euclid's proof in some detail as a point of reference to understand which elements or steps of it have been changed or deleted in modern proofs. Some analysis of the structure and methods of Euclid's proof is also needed. The following short summary of proof methods is (for the convenience of modern readers) based on one simplification compared to the original formulation. As will become clear from the text that follows, the summary replaces Euclid's notion of "composite" number ("which is measured by some

Take the least number DE measured by A, B, and C. Add the unit DF to DE.

Then EF is either prime or not.

First, let it be prime. Then the prime numbers A, B, C, and EF have been found, which are more than A, B, and C.

Next, let EF not be prime. Therefore it is measured by some prime number [VII, 31].[20]

Let it be measured by the prime number G.

I say that G is not the same with any of the numbers A, B, and C.

---

[16]There is an old historical discussion among historians of Greek mathematics (Wilbur Knorr, David Fowler, etc.) that continues today about the relation between constructive and deductive principles in the mathematics of the time and the influence of philosophy on that relationship.

[17]This algorithm was certainly mathematically deeper than ECO and ECL, and Donald Knuth called it the "granddaddy of all algorithms" (Knuth 1981: 318). The Euclidean algorithm is not needed for Euclid's proof in IX, 20, although finding the greatest common divisor is closely related to finding the least common multiple LCM, which is assumed in IX, 20. The reason is that the LCM is trivial for prime numbers. However, as remarked earlier, Graham/Knuth/Patashnik (1989: 108) use the Euclidean algorithm to construct an infinitude of primes recursively.

[18]The essentially arithmetic content of the "number-theoretic books VII–IX" in the *Elements* is stressed by Knorr (1976). Vitrac (1994, 277) emphasizes differences in detail between the constructions of book I in the *Elements* and the constructions in the number-theoretic books.

[19]David E. Joyce uses Heath's edition from 1908 for his online edition at the website of Clark University, from which we have quoted here: http://aleph0.clarku.edu/~djoyce/java/elements/elements.html. Figures have been added in Heath/Joyce and in some other editions for explanation. Heath/Joyce follow closely the Greek version in J. L. Heiberg's edition of 1883–1885, replacing the Greek letters in the same alphabetic order. Cf. page 271 of the bilingual online edition by Richard Fitzpatrick at http://farside.ph.utexas.edu/euclid/elements.pdf. Although it does not include genuine geometric content, the figure is useful, because the text speaks of the least common multiple DE, and one would expect the unit to be called EF, not DF. The picture explains that. I thank June Barrow-Green for this observation.

[20]Vitrac 1994, 444, refers here to VII, 32 instead, which is a simple conclusion of VII, 31, but does not contain the construction of the divisor. Explicit reference to previous theorems (as explanation) is usually not contained in Euclid's text and is not added in Heiberg's Greek edition.

If possible, let it be so. Now A, B, and C measure DE, therefore G also measures DE. But it also measures EF. Therefore G, being a number, measures the remainder, the unit DF, which is absurd.

Therefore G is not the same with any one of the numbers A, B, and C. And by hypothesis, it is prime. Therefore the prime numbers A, B, C, and G have been found, which are more than the assigned multitude of A, B, and C.

Therefore, *prime numbers are more than any assigned multitude of prime numbers.*

<div align="right">Q. E. D.</div>

Paraphrasing of the Proof:

Given any finite set S of primes, one considers (constructs) their product P [Euclid uses for some not fully clear reason[21] the LCM instead of the product; R. S.-S.] and adds the unit 1. If $P + 1$ is a prime, one has found an additional prime, which means a prime that is not in the original set S (L1). If $P + 1$ is not a prime, it is divided by a prime (T). This latter prime cannot, however, be one of the primes in the original finite set either, because in this case it could not divide $P + 1$ (L2). Therefore we have also in this case found an additional prime that does not lie in the original set S of prime numbers assumed.

This is a paraphrasing of Euclid's proof (or of what has come down to us in handwritten copies and in various Arabic, Latin, and modern translations of the Greek original), which follows closely Heiberg's Greek edition of Euclid's text, although partly using modern vocabulary (set).

It is remarkable that the use of lemma L1 is not expressly mentioned or argued within Euclid's proof. It is simpler than L2 but in its content closely related to it. It shows that $P + 1$ (regardless of whether P is the LCM or the product of all given primes) is bigger than any prime in S and therefore cannot be one of these primes. However, L1 is often emphasized in modern presentations of the theorem, for instance in the one by Dirichlet to be mentioned in the text that follows. L1 is necessary for Euclid's proof to show that $P + 1$ lies outside S.

Is Euclid's Proof One by Contradiction (in Addition to Being One by Construction)?

An indirect proof of the infinitude of primes has to be based on the (false) assumption that the given finite set of primes is "complete." Now the question arises as to whether Euclid's proof could at the same time—in addition to being weakly constructive—be interpreted as an indirect one, by assuming that S comprises all primes and by disproving this assumption. In other words, could one conclude MCL immediately from the original proof ECO? After all, it seems at first glance that S could well have been assumed by Euclid to be both finite *and* complete. Thus the latter assumption would have been disproved by ECO and Euclid would have produced a contradiction.

H&W make much of the fact that Euclid does not expressly say that the proof uses the assumption that S is complete. However Euclid (and many other mathematicians with him, even today) does not always fully explain what he does and intends. We conclude, as a matter of course, that in Euclid the three numbers A, B, C stand for an arbitrary finite set of primes, the cardinality of which we would today describe by the indeterminate number n. We take for granted that Euclid saw that the new prime he had "constructed" was $\leq P + 1$. So why not assume that he had thoughts about the completeness of S as well?

We do, however, have a compelling and very simple argument to discard the hypothesis that Euclid assumed S to be complete.[22] This argument lies, of course, in the proof itself. In it Euclid concludes with the help of L1 that $P + 1$ lies outside S. But he then considers as one of two possible cases the one that $P + 1$ could be prime. But this is a conclusion that Euclid never would have drawn under the assumption that S was complete. To perform a *reductio ad absurdum* requires, of course, concluding correctly from a hypothesis that one wants to disprove. Indeed, Dirichlet in his (indirect) proof (see later), which assumes the completeness of S, discards immediately the possibility of $P + 1$ being prime.

Note that Euclid's proof, although not *globally* assuming the completeness of primes, nevertheless uses *locally* "indirect arguments" as steps in the proof. And this use occurs even on two levels, one open and one concealed. In the middle of the proof, the assumption of $P + 1$ being divisible by a prime in S is refuted with the help of lemma L2. And on a more hidden level: The proof of theorem T "Any product of numbers (none of them the unit) is divisible by a prime number" uses an indirect argument, which in Heath's English edition appears within the proof of VII, 31 as:

Thus, if the investigation be continued in this way, some prime number will be found which will measure the number before it, which will also measure A. For, if it is not found, an infinite series of numbers will measure the number A, each of which is less than the other: which is impossible in numbers.

Vitrac in his French edition of the *Elements* emphasizes that we have an infinity argument here that is reminiscent of the "descent infinite which has been made famous by Fermat." (Vitrac 1994: 341). To be sure, this "hidden" indirect argument is of no immediate concern when deciding whether Euclid's proof in IX, 20 is "indirect," because theorem T (= VII, 31) is accepted in the proof as a mathematical fact. However, the use of T nevertheless contributes to the "feeling" that we have to do with indirect arguments, because theorem T is obviously partly responsible for the "weak" constructiveness of IX, 20. This is even more true of the unconcealed indirect argument mentioned before. Because it seems difficult to rewrite these two indirect arguments in the form of a direct conclusion (which would imply that the indirect argument is merely

---

[21]The reason may be again the emphasis of properties over constructions.
[22]H&W mention this as well as an additional argument (p. 46).

superficial),[23] we are left with the result that the proof IX, 20 contains locally genuinely indirect arguments, but globally it is not indirect because it does not assume the completeness of the given finite set of primes.

Insinuating Contradiction and (Almost) Ignoring Construction: In Particular in G. H. Hardy

H&W note that Euclid's proof is misinterpreted by "no less a number theorist" than G. H. Hardy (1877–1947) in his *Course of Pure Mathematics* (1908). Because they do not quote Hardy in detail and because this fascinating example serves my further argument, I present here the full passage from the original of 1908, which is repeated in subsequent editions up to the sixth of 1933:[24]

> Euclid's proof is as follows. If there are only a finite number of primes let them be l, 2, 3, 5, 7, 11, … N. Consider the number l + (l. 2. 3. 5. 7. 11 … N). This number is evidently not divisible by any of 2, 3, 5, … N, since the remainder when it is divided by any of these numbers is 1. It is therefore not divisible by any prime save l, **and is therefore itself prime** [my emphasis, R. S.-S.], which is contrary to our hypothesis (Hardy 1908: 122/123).

Hardy, unlike Euclid, uses primorials in his proof, that is, the products $p_n\#$ of the first n primes. We will discuss this change (maybe even distortion) further below. In addition, Hardy assumes these n primes to form a *complete* set of primes ("there are only"). Surprisingly, Hardy considers the number 1 as a prime. But what is really striking is the claim that $P + 1$ "is therefore itself prime." An inexperienced student can easily jump to the conclusion that $P + 1$ must indeed be prime whatever prime number N one starts with.

This is, of course, not the case, as the famous and simplest counterexample, $p_6\# + 1 = 2 \times 3 \times 5 \times 7 \times 11 \times 13 + 1 = 30{,}031 = 59 \times 509$, shows. Of course, we should not assume that the great English number theorist was not aware of the counterexample. To the contrary, he might well have been "too aware" of it, thus committing a "pedagogical error." Indeed, the conclusion $P + 1$ "is therefore itself prime" might be convincing for the beginner for two reasons: it is "confirmed" by the first five "Euclid numbers"[25] up to $p_5\# + 1 = 2311$, and it is supported by the intuitive, if misled, feeling that there cannot be a divisor of $P + 1$ between N and $P + 1$. However, the conclusion "is therefore itself prime" will be considered "absurd" by the mathematically educated and thus completes for him the *reductio ad absurdum*. Because one can correctly deduce any statement from a wrong one, the conclusion that $P + 1$ must be prime is one that should not be too surprising for the educated mathematician.

Hardy does not bother to extend the last sentence of his proof with a remark such as "…which is contrary to our hypothesis, and the latter has therefore to be discarded." Hardy does not provide a definition of a prime number in his 1908 book either, which would be important to know for reconstructing his argument. He apparently concluded along these lines: The constructed number, which I will again call $P + 1$, is not divisible by "any prime" in S (L2). $P + 1$ can therefore—because of the completeness of S (assumption A) and theorem T—not be composite and must be prime. But then it follows from the completeness of S and from L2 that $P + 1$ cannot be divisible by itself and thus cannot be prime. This is a contradiction and it follows "non-A" (symbolically Ā), which is the original claim one wants to prove.

Symbolically Hardy's indirect modern proof of the infinitude of primes MIP-H1908 can thus be written in the following way, although he does not refer to the second step of the argument in his verbal formulation:

$$(H\,1908)\ [A \xrightarrow{L2+T} (P+1)prime] \wedge [A \xrightarrow{L2} (P+1)\ \neg prime] \rightarrow \neg A)$$

Compared to Euclid, Hardy—in addition to assuming the completeness of S (assumption A)—exchanges the steps of the proof, and he does not use lemma L1 at all. Instead of (similar to a hypothetical indirect proof in Euclid) first considering the case of $P + 1$ being prime and recognizing it as a contradiction to A because of L1, Hardy considers first $P + 1$ being composite and discards it as a contradiction to A because of L2. By exchanging the steps of proof, he is left with the conclusion that $P + 1$ is prime under the assumption of A, which is then finally recognized as a contradiction.

One could argue that Hardy liked to play the logical game to the end and to the extreme, namely to conclude the absurd statement that $P + 1$ is "prime" under the (wrong) assumption of A. He seems, however, less interested in the constructive conclusion of Euclid's proof, which considers the divisibility of $P + 1$ as a composite number. Thus, Hardy distorts Euclid's proof ECO in several respects, assuming the completeness of S, using primorials, and changing the steps in Euclid's arguments. And one realizes that the order of steps and the wording matter when one analyses the various forms of the presentation of Euclid's proof.

Compare now Hardy's proof with the one by Dirichlet in the posthumous edition of his *Number Theory* of 1863.[26] Starting with the maximum prime p in the finite set S (which is assumed to be complete) and (like Hardy after him) using primorials, Dirichlet continues:

> Each number greater than p must be composite and hence divisible by at least one of these primes. But it is very easy to construct a number greater than p and not divisible by any of the primes: just construct the product

---

[23]Note that any constructive proof can be rewritten as an indirect proof by assuming the opposite. But it should be superficial to acknowledge the latter as a genuine indirect proof.

[24]I have checked the sixth edition of 1933 and can confirm that Hardy changes the passage for the first time in the seventh edition of 1938, which he calls "revised and re-set." See below.

[25]This is historically clearly a misnomer. "Euclid number" as used in Graham/Knuth/Patashnik (1989) is, however, somewhat closer to Euclid.

[26]Dirichlet 1863 and 1999, pp. 15–16, edited by R. Dedekind 1863 and translated into English by J. Stillwell in 1999. This is mentioned without discussion in Hardy (2013), which contains some additional remarks and slight corrections to H&W (2009).

of the primes from 2 to p and add l (Dirichlet 1863/1999, 10).

From lemma L2, Dirichlet then concludes a contradiction to the conclusion that $P + 1$ is composite. A comparison with Euclid is—once again—difficult because of the assumption of the completeness of S in Dirichlet. However, one can say that, in a sense, Dirichlet is more faithful than Hardy to Euclid's original proof, because—like Euclid—he first (implicitly) considers $P + 1$ to be prime, but then discards this possibility under the assumption of A, because $P + 1$ is greater than p. Dirichlet thus even uses expressly lemma L1, which is hidden in Euclid. That lemma implies that because of the multiplication that produces $P + 1$, the latter number must lie outside the original set S.

Symbolically one could write Dirichlet's indirect proof MIP-D1863 in the following way:

$$(D\,1863)\ [A \xrightarrow{\;L1\;} (P+1)composite] \wedge [A \xrightarrow{\;L2\;} (P+1)$$
$$\neg composite] \rightarrow \neg A$$

Dirichlet's proof seems to me one of the shortest and most elegant proofs of MCL, that is, the infinitude of primes. It uses—as any other proof that connects to Euclid—the main constructive idea, namely to construct and investigate the number $P + 1$. However Dirichlet—like Hardy after him—is more interested in the fact of infinitude (MCL) than in the algorithmic problem of finding an additional prime in finitely many steps. Given that Dirichlet himself had proven much sharper results on the location of the infinitely many primes[27] than can be concluded from ECO, this lack of interest is not surprising.

Why is it that Dirichlet's indirect proof appears more natural and less logically confusing for beginners than Hardy's of 1908? One reason is that the (absurd) conclusion that $P + 1$ should be composite is refuted by a much simpler numerical counterexample ($2 \times 3 + 1 = 7$). The stronger pedagogic appeal in Dirichlet seems to me also a result of Dirichlet following closer to Euclid's well-known proof and using—unlike Hardy—the very intuitive lemma L1. One could even argue that Dirichlet is more pedagogic than Euclid, that is, Dirichlet shows the mathematical instruments that he uses more clearly than Euclid. Indeed, one should not look at modern presentations of Euclid only under the perspective of distortion but also under the point of view of logical clarification.

Apparently, in the end Hardy himself recognized his "pedagogical error." In the seventh edition of 1938 of his book (in which he also no longer considered 1 to be a prime) Hardy wrote:

There are however, as was first shown by Euclid, infinitely many primes. Euclid's proof is as follows. Let 2, 3, 5, ... p_N be all the primes up to p_N, and let $P = (2.3.5... p_N) + 1$. Then P is not divisible by any of 2, 3, 5, ..., p_N. Hence either P is prime, or P is divisible by a prime p between p_N

and P. In either case there is a prime greater than p_N and so an infinity of primes (Hardy 1938, p. 125).

Despite this dramatic change in the later editions of his *Course of Pure Mathematics* and his return to Euclid's original proof, Hardy continued to misunderstand the latter as one by contradiction. Hardy gave Euclid's proof a prominent place in his famous *Apology* of 1940, where he presented basically the same newly corrected version of the proof from his *Course*:

The proof is by *reductio ad absurdum*, and *reductio ad absurdum*, which Euclid loved so much, is one of a mathematician's finest weapons. It is a far finer gambit than any chess gambit: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game (Hardy 1940: 92).

But at the same time and in the same book of 1940, Hardy explained that Euclid's theorem is just a fundamental result about the infinitude of primes and that it was limited in its possible applications.

H&W rightly find that presentations such as the one by Hardy in 1908, which insinuate a proof by contradiction and neglect construction, lack "simplicity," and that Euclid's proof was "simpler." I believe that Euclid's proof is, indeed, "simpler" because it is better pedagogically and does not resort to wording that makes the notion of "primality" contingent on absurd assumptions. However, if they mean that Euclid's proof is simpler because (!) it is constructive, I disagree. On the contrary, one could argue that a proof purely by contradiction is simpler because it has less content.[28]

Least Common Multiple, Primorials, Factorials: The Logical Equivalence of the Proofs

Let's return to the question I asked at the beginning: is the "spirit" and "content" of Euclid's proof preserved in modern presentations that use primorials or even factorials, although Euclid uses products of any finite number of given primes? Does this actually affect the presentation of the "substance" of Euclid's result?

H&W rightly criticize remarks such as this one by Tobias Dantzig (1884–1956), the father of the better-known mathematician George Dantzig:

In this proof Euclid introduces for the first time in history what we call today factorial numbers (Dantzig, quoted by Hardy/Woodgold 2009: 47).

This was even less faithful to Euclid than the use of primorials. Of course an historian has to reject such wrong attributions to original sources when they come as bluntly and as unnuanced as in Dantzig's book. A similar remark applies to Dickson's book (1919), which the author calls himself a "History of Number Theory."[29]

Now we find another use of factorials in an allusion to Euclid's proof in David Hilbert's talk "Über das

---

[27]Namely, the existence of an infinitude of primes in arithmetic sequences.

[28]For example, as described earlier, it is simpler to prove the *actual* infinitude of all primes, because of the additional indirect argument that leads to MCL, than to construct the *potential* infinitude of particular sets of primes with the help of ECO.

[29]Dickson (1919: 413) says mistakenly that he quotes directly from Heiberg's Greek edition when he alleges that Euclid used primorials. Thanks go to June Barrow-Green for alerting me to Dickson.

Unendliche" (On the Infinite) in the context of a discussion of "finitism." This talk may have influenced later presentations of Euclid's theorem.[30] Indeed, in the English translation by the remarkable German mathematician and musician Stefan Bauer-Mengelberg,[31] we read:

By means of Euclid's well-known procedure we can, completely within the framework of the attitude we have adopted [i.e., the "finitist" in Hilbert's sense as described in his quote that follows, R. S.-S.], prove the theorem that between $p + 1$ and $p! + 1$ there certainly exists a new prime number. This proposition itself, moreover, is completely in conformity with our finitist attitude. For "there exists" here serves merely to abbreviate the proposition:

Certainly $p + 1$ or $p + 2$ or $p + 3$ or … or $p! + 1$ is a prime number.

But let us go on. Obviously, to say there exists a prime number that (1) is $> p$ and (2) is at the same time $\le p! + 1$ would amount to the same thing, and this leads us to formulate a proposition that expresses only a part of Euclid's assertion, namely: there exists a prime number that is $> p$. So far as content is concerned, this is a much weaker assertion, stating only a part of Euclid's proposition; nevertheless, no matter how harmless the transition appears to be, there is a leap into the transfinite when this partial proposition, taken out of the context above, is stated as an independent assertion.

How can that be? We have here an existential proposition with "there exists." To be sure, we already had one in Euclid's theorem. But the latter, with its "there exists," was, as I have already said, merely another, shorter expression for

"$p + 1$ or $p + 2$ or $p + 3$ or . . . or $p! + 1$ is a prime number,"

just as, instead of saying: This piece of chalk is red or that piece of chalk is red or … or the piece of chalk over there is red, I say more briefly: Among these pieces of chalk there exists a red one. An assertion of this kind, that in a finite totality "there exists" an object having a certain property, is completely in conformity with our finitist attitude. On the other hand, the expression

"$p + 1$ or $p + 2$ or $p + 3$ or … *ad infinitum* is a prime number"

is, as it were, an infinite logical product,[32] and such a passage to the infinite is no more permitted without special investigation and perhaps certain precautionary measures than the passage from a finite to an infinite product in analysis, and initially it has no meaning at all.

In general, from the finitist point of view an existential proposition of the form "There exists a number having this or that property" has meaning only as a *partial proposition,* that is, as part of a proposition that is more precisely determined but whose exact content is unessential for many applications (Hilbert 1925: 377–378).

Now Hilbert is known not to have been very history-minded, and he was usually not very concerned about accuracy in his allusions to the history of his discipline. In the example under discussion here, Hilbert does not give any details about Euclid's proof either. However, his claim that Euclid reached two different conclusions, namely "a much weaker assertion" about infinitely many primes and a stronger one specifying a finite set where the additional primes exist is in good agreement with what I argued earlier.

It remains to be seen whether Hilbert provided an accurate description of the conclusions that could be drawn from "Euclid's well-known procedure," in particular, how much the use of factorials changed Euclid's original argument. It is of course obvious that $p!$ is divisible by all natural numbers $< p$, in particular by all primes $< p$. If $p! + 1$ is not prime, it is divisible by a prime according Euclid VII, 31 (theorem T). This prime must be smaller than the usually very big number $p!$, but bigger than p, because otherwise the classic contradiction from Euclid's theorem of a division with a remainder < 1 occurs. Therefore there must be a prime between $p + 1$ and $p!$. Thus the main idea is indeed analogous to Euclid; Hilbert's prime number p can be considered as the biggest prime number in Euclid's set S. Hilbert has thus shown, with the same argument as Euclid, that there exists a prime number bigger than any given prime number p and smaller than an upper limit depending on a concrete natural number that can be calculated from p.

Whereas Euclid proves that there exists a larger set (!) of prime numbers than any finite set of prime numbers, Hilbert proves that there exists a larger prime number than any given prime number. Both Euclid and Hilbert provide concrete boundaries for the new set or the new number. Both proofs are logically equivalent, because any finite set of numbers has a largest element and conversely one can find (for instance with the well-known sieve of Eratosthenes) to a given p all prime numbers smaller than it and thus one can find any of the upper limits for an additional prime.

The Obsolescence of Euclid's Proof and Some Concluding Remarks

I find three main historical reasons for the misrepresentation of Euclid's proof as indirect.

The first reason is historical fashion. Indeed, the reference of modern, logically minded mathematicians (from the second half of the 19th and a good part of the 20th centuries) to the old ideal of Euclid's *Elements* after several "constructive centuries" (Descartes through Euler) is not coincidental. They know that Greek mathematicians used

---

[30]However, Meštrović (2012: 9) mentions an article by H. Brocard of 1915 that seems to indicate that Charles Hermite already used factorials for the presentation of Euclid's proof.

[31]Stefan Bauer-Mengelberg (1927–1996) was a German-born mathematician who worked at I.B.M. and had a simultaneous career as a conductor. He served for a while as assistant conductor to Leonard Bernstein at the New York Philharmonic. Cf. his obituary in *The New York Times*, published October 28, 1996, also accessible online.

[32]Editor Jan van Heijenoort adds here the following footnote: ''It is rather a logical sum or disjunction. In the version published in Grundlagen der Geometrie (1930), 'logisches Produkt' is replaced by 'Oder-Verknüpfung'.''

indirect proofs widely and thus they read into Euclid what they want to stress. The modern mathematicians are helped by the fact that Euclid himself in the *Elements* downplayed the constructive aspects of the proofs or at least did not mention them in the statements of many theorems. This in particular is the case for Euclid's theorem of the infinitude of primes, which does not mention in its claim the mainly constructive nature of the proof. From the 1960s, mathematicians such as D. Knuth and G. Pólya, and philosophers such as I. Lakatos, have contributed with their work to a renaissance of constructive and algorithmic methodologies in mathematics. This was of course gradually supported by the rise of data technology and computer science, which left its mark in prime number theory as well, the prime numbers finding unexpected applications in coding theory and so forth, for example, in the banking business.

The second historical reason for misinterpreting Euclid's proof as one by contradiction is the weak form of constructiveness that does not provide an effective procedure to find individual primes: Euclid constructed a finite set of numbers within which the existence of an additional prime was guaranteed; however this additional prime (or primes) can only be found by testing the elements of this finite set consecutively—and apparently leaving the mathematician without information about a preferential order—for primality. David Hilbert referred to this weaker notion of construction, which is exhibited by Euclid's proof, in the context of his "finitist" approach to the foundations of mathematics in the 1920s and 1930s. The weak constructiveness is also connected to a partial and "local" use of indirect arguments within the proof. One gets the impression that some modern mathematicians looking at Euclid's weak construction compare it with some modern proofs by contradiction, not of Euclid's theorem, but, for instance, in set theory, which indeed are not constructive in any sense.

There is in my opinion, however, a **third historical reason**, no less important, for many modern authors to insinuate an indirect proof of Euclid's theorem and to downplay its constructiveness. On this historical reason I will elaborate a bit in these concluding remarks. Although the infinitude of primes is still a basic and important fact in number theory, the insight of modern mathematicians into the distribution of primes goes far beyond what is indicated in Euclid's constructive proof, which is therefore mathematically obsolete, the entire theorem being a very trivial one. Indeed, Euclid's proof, both in its original and in its modern forms, offers little information about the distribution of primes. According to Euclid's original proof, additional primes can also be sought below the biggest given prime, as the simple examples $3 \times 5 + 1 = 16$ and $2 \times 7 + 1 = 15$ show. Euclid's original argument allows, for example, for any prime $p > 2$ to conclude the existence of an additional prime $< 2p$. However, Euclid's proof does not rule out that this prime should be smaller than p: it was only shown much later by P. L. Chebysev (1850),

with stronger analytical methods, that in any case there always exists a prime between p and $2p$.[33] That there can at the same time exist additional primes smaller than p and between p and $2p$ is shown for p = 7 and the example $2 \times 7 + 1 = 15$. Thus Chebysev's result is much sharper than Euclid's in this case.

Hardy and Wright ([1938](#)) reflect on the conclusions that can be drawn from Euclid's proof for $\pi(x)$, the number of primes below x. They present ECO very close to Euclid's spirit as in Hardy's corrected version discussed earlier. Then they present, based on the use of primorials, upper limits for the increase of p_n which is the value of the nth prime, and lower limits for $\pi(x)$. Hardy's and Wright's discussion shows that the two mathematicians were well aware of the constructive side of Euclid's proof, but that they deemed its potential to be "absurdly weak"[34] compared to what was already known at the time about the distribution of primes, based for instance on the (logarithmic) prime number theorem (proven in 1896 by Hadamard and de la Vallée Poussin) and other more refined results. This provides further confirmation of the reason for G. H. Hardy's mistaken claim that Euclid's argument was mainly by contradiction. Euclid's theorem apparently was to him only of a general "philosophical" interest, not a concrete help for finding prime numbers.

One gains still another perspective on the constructive/ algorithmic side of Euclid's theorem when looking, for instance, at the book by Boolos, Burgess, and Jeffrey, which uses factorials:

7.10 Example (The next prime). Let $f(x)=$ the least y such that $x < y$ and y is prime. The relation $x < y$ & y is prime is primitive recursive, using Example 7.5. Hence the function f is recursive by the preceding proposition. There is a theorem in Euclid's *Elements* that tells us that for any given number x there exists a prime $y > x$, from which we know that our function f is total. But actually, the proof in Euclid shows that there is a prime $y > x$ with $y \leq x! + 1$. Since the factorial function is primitive recursive, the Corollary 7.8 applies to show that f is actually *primitive* recursive (Boolos, Burgess, and Jeffrey [1980](#): 79).

Thus modern work on logic and computability apparently has drawn—at least on a theoretical level—some inspiration from the constructive part of Euclid's proof, here in Boolos et al, in connection with the theory of primitive recursive relations and assuming the use of factorials.[35]

But Boolos et al do not share the concerns of some "heterodox mathematicians who reject certain principles of logic." Referring to Hilbert's "finitism," they point to the power of modern, analytical mathematics, interestingly enough in a number-theoretic context:

On the plane of mathematical practice, Hilbert insisted, a detour through the "ideal" is often the shortest route to a "contentful" result. (For example, Chebyshev's theorem that there is a prime between any number and its double

---

[33]Hardy and Wright ([1938](#): 13). See the later remark quoted from Boolos et al ([1980](#): 238). The theorem is valid even for arbitrary n, not necessarily prime.

[34]Cf. a similar remark in Meštrović ([2012](#): 30), who calls it a ''horrible bound.''

[35]Cf. also Graham, Knuth, and Patashnik ([1989](#)), as quoted earlier, and the note by Mullin (1963) who constructs with the ''Euclidean idea'' of the number $P + 1$ a sequence of primes that never generates any prime twice. Mullin refers directly back to E. L. Post's theory of recursively enumerable sets of positive integers and their decision problems (1944). Sequences of primes have certainly played an increasing role in mathematical logic, for instance in the method of Gödelization.

was proved not in some "finitistic," "constructive," directly computational way, but by an argument involving applying calculus to functions whose arguments and values are imaginary numbers.) (Boolos, Burgess, and Jeffrey 1980: 238).

As is well known, Hilbert's version of "finitism" was, in reality, a defense strategy to save classic, infinitist mathematics. Also, G. H. Hardy was not willing to renounce the use of modern analysis in number theory when in his *A Mathematician's Apology* he said about Euclid's theorem:

> The proof can be rearranged to avoid a *reductio*, and logicians of some schools would prefer it should be (Hardy 1940: 94).

It is not fully clear what Hardy meant by rearrangement. I suspect that he did not have in mind a simple rearrangement of Euclid's proof but rather a strict and formalized rewriting of the proof in the language of constructive mathematics. In any case, this (like Hilbert's presentation) says less about what Euclid's proof was (whether by contradiction, or construction, or both) than about what Hardy wanted it to have been, determined by his own research interests (which were certainly not directed toward constructivism) and by the spirit of the time.

To Hilbert, the constructive side of Euclid's proof was probably of philosophical rather than mathematical interest, and was not of practical concern. Since Hilbert's times, much better proofs and methods have been devised, which make it fairly easy to find prime numbers that have almost any characteristic one wants them to have. Many modern searches for higher primes use special numbers such as Mersenne numbers, for which there exist strong criteria of primality.[36]

The examples of Dirichlet, Hardy, Hilbert, and other leading mathematicians of the 19th and 20th centuries reading Euclid's theorem, corroborate a recurring theme in this article, namely that modern views of the history of mathematics are often colored by current research interests.

Faculty of Engineering and Science
University of Agder
Gimlemoen, Postboks 422
4604, Kristiansand S
Norway
e-mail: reinhard.siegmund-schultze@uia.no

REFERENCES

Barrow-Green, J., 2006: ''Much necessary for all sortes of men'': 450 years of Euclid's Elements in English, *BSHM Bulletin* 21, 2–25.

Boolos, G., Burgess, J. P., and Jeffrey, R. C., 1980: *Computability and Logic,* Cambridge: Cambridge University Press.

Dirichlet, P. G. L., 1863: *Lectures on Number Theory (supplements by R. Dedekind)*; [Publication of Dedekind's posthumous edition of 1863 of Dirichlet's lectures with later supplements I-IX by Dirichlet, translated by J. Stillwell], AMS, 1999.

Dickson, L. E., 1919: *History of the Theory of Numbers*, vol. 1, Divisibility and Primality. Washington: The Carnegie Institution.

Euclid, 1908, 1926: *The Thirteen Books of The Elements, translated with introduction and commentary by Sir Thomas Heath*, New York: Dover 1956.

Euclide d'Alexandrie, 1990–2001: *Les Éléments, traduit du texte de Heiberg, Traduction et commentaires par Bernard Vitrac*; 4 volumes, Paris : Presses Universitaires de France.

Graham, Ronald L., Donald E. Knuth, and Oren Patashnik, 1989: *Concrete Mathematics: A Foundation for Computer Science, Reading, etc*. Boston: Addison-Wesley.

Hardy, M., 2013: Three Thoughts on ''Prime Simplicity,'' *The Mathematical Intelligencer* 35(1), 2.

Hardy, M., and Woodgold, C., 2009: Prime Simplicity, *Mathematical Intelligencer* 31(4), 44–52

Hardy, G. H., 1908: *A Course of Pure Mathematics*, Cambridge: Cambridge University Press.

Hardy, G. H., 1938: *A Course of Pure Mathematics*, 7th edition, Cambridge: Cambridge University Press.

Hardy, G. H., 1940: *A Mathematician's Apology*, Cambridge: Cambridge University Press.

Hardy, G. H., and E. M. Wright, 1938: *An Introduction to Number Theory*, Oxford, Oxford University Press.

Heijenoort, J. v. (ed.), 1967: *From Frege to Gödel*, Cambridge: Harvard University Press.

Hilbert, D., 1925: Über das Unendliche, *Mathematische Annalen* 95 (1926), 172, quoted from the English translation by Stefan Bauer-Mengelberg in Heijenoort (1967), 367–392.

Knorr, W., 1975: *The Evolution of the Euclidean Elements*, Dordrecht: Reidel.

Knorr, W., 1976: Problems in the Interpretation of Greek Number Theory: Euclid and the ''Fundamental Theorem of Arithmetic,'' *Studies in Hist. and Philosophy of Science* Part A 7 (4), 353–368.

Knuth, D. E., 1981: *The Art of Computer Programming*, vol. 2, *Seminumerical Algorithms*, 2nd edition: Reading: Addison-Wesley.

Meštrović, Romeo, 2012: Euclid's theorem on the infinitude of primes: A historical survey of its proofs (300 BC to 2012), http://arxiv.org/abs/1202.3670v2, 5 June 2012, 66p.

Mullins, A. A., 1963: Recursive Function Theory (A Modern Look at a Euclidean Idea), *Bulletin AMS* 69, 737.

Nikolić, M., 1974: The Relation between Eudoxos' Theory of Proportions and Dedekind's Theory of Cuts, In: Cohen, R. S., J. J. Stachel, and M. W. Wartofsky (eds.), *For Dirk Struik: Scientific, Historical and Political Essays in Honor of Dirk J. Struik*, Dordrecht: D. Reidel, 225–243.

---

[36]Cf., for example, Ribenboim (1996).

Ribenboim, P., 1996: *The New Book of Prime Number Records*, New York: Springer.

Unguru, S., 1975: On the Need to Rewrite the History of Greek Mathematics, *Archive for the History of Exact Sciences* 15, 67–114.

Vitrac, B., 1994: *Euclide d'Alexandrie, 1990–2001*, volume 2.

Weil, A., 1978: History of Mathematics: Why and How? *Proceedings of the International Congress of Mathematicians, Helsinki* 1, 227–236.

Mathematics Pure or Applied

Martin Zerner reports that a few years ago he was scanning a list of mathematical specialties put out by the Conseil National des Recherches Scientifiques, and noticed that one of the categories was

Applications of Pure Mathematics.

With Zerner, we may furrow our brows. Is it an oxymoron? If not, maybe the CNRS ought to have allowed also a category

Purification of Applied Mathematics

–which, if it means anything, surely means something rather inglorious. Like removing from some discipline the stigma of applicability (as if there were such a stigma). Joseph Keller's dismissal of all such anxieties is

Pure mathematics is a subfield of applied mathematics.

Chandler Davis