Chapter 6
The Fundamental Theorem of Arithmetic

The Fundamental Theorem of Arithmetic (FTA) states that every integer greater than 1 has a factorization into primes that is unique up to the order of the factors. The theorem is often credited to Euclid, but was apparently first stated in that generality by Gauss.[1] Note that the statement has two parts: First, every integer greater than 1 *has* a factorization into primes; second, any two factorizations of an integer greater than 1 into primes must be identical except for the order of the factors. The proofs of each of those parts will thus be considered separately.

The first part is the subject of propositions VII,31 and VII,32 of the *Elements*. Proposition VII,31 states that any composite number (that is, any number that has a proper divisor other than one) is divisible by some prime. Having established that, Euclid then immediately concludes in proposition VII,32 that any number greater than 1 is either prime or is divisible by some prime.

Euclid's proof of VII,31: Let A be a composite number. By definition, A has a proper divisor B other than one. If B is prime, we are done. If not, B has a proper divisor C other than one, and then C is a proper divisor of A. If C is prime, we are done. Otherwise, C has a proper divisor other than one. Continuing in this fashion, one must eventually obtain a prime divisor of A, since otherwise there would be an infinite sequence of divisors $B, C, \ldots$ of A, each smaller than the one before, which is impossible.

Second proof (of VII,31 and VII,32 together): By complete induction on the integer $A > 1$. Suppose every integer greater than 1 and less than A is divisible by some prime. Consider A. If A is prime, we are done. Otherwise, $A = BC$ with $1 < B, C < A$. By the inductive hypothesis, B is divisible by some prime, and that prime divides A.

[1] In the *Disquisitiones Arithmeticae* (Gauss 1801, Bd. I, p. 15). (See Collison 1980, p. 98.) However, the result was certainly known, if not explicitly stated, beforehand. For example, Euler used it implicitly in his 1737 proof of the infinitude of the primes (via the divergence of the harmonic series).

J.W. Dawson, Jr., *Why Prove it Again?*, DOI 10.1007/978-3-319-17368-9_6

Repeated application of VII,32 then establishes the existence of a prime factorization for any integer greater than 1.

Note that the first of the proofs above is a *reductio*, while the second is a direct proof that explicitly uses induction. Euclid takes for granted that there cannot be an infinite strictly decreasing sequence of positive integers — a statement that nowadays would be deemed to require proof. The most direct proof is by means of the well-ordering principle, which is equivalent to induction (or complete induction). Indeed, the statement in question is itself logically equivalent to induction.

There is little more development of the first part of the FTA, since the second argument above is so simple. (Some further remarks about the first part will, however, be made at the end of the chapter). Consider then the second part of the FTA.

A corollary of the FTA is Euclid's Lemma, which asserts that if a prime divides a product it must divide one of the factors. A strong form of Euclid's Lemma, but restricted to products of just two factors, was stated by Euclid as proposition VII,30 of the *Elements*. A consequence of that result is Euclid's proposition IX,14 ('If a number be the least that is measured by [three distinct] prime numbers, it will not be measured by any other prime number except those originally measuring it.').

Euclid did not consider products of more than three primes, nor products involving repeated factors. However, his proof of IX,14 can be applied to conclude that the representation of a number as a product of *distinct* primes is unique except for the order of the factors. To extend to products involving repeated factors one can apply proposition IX,13 of the *Elements* (which states that the only divisors of p^k are the numbers $1, p, p^2, \ldots, p^k$) in combination with VII,30. Alternatively, one can argue, as Gauss later did, that if a prime p appears to the power j in one factorization of a number n and to the power k in another, with $j \leqslant k$, then dividing by p^j will yield two factorizations of another number, at most one of which involves the factor p. Applying VII,30 repeatedly then shows that p must in fact occur in neither, so $j = k$. Any other repeated prime factors may be similarly eliminated, so only products of distinct prime factors need be considered.

Consequently, Euclid's Lemma also implies the second part of the FTA, and it is useful to distinguish proofs of the second part of the FTA that do *not* first prove Euclid's Lemma from those that do. Proofs of the second part of the FTA may also be distinguished according to what extent (if any) they use mathematical induction, whether they are direct or indirect, whether they invoke the concepts of least common multiple or greatest common divisor, and whether they employ the division algorithm, the Euclidean algorithm, or neither.

6.1 Direct proofs of Euclid's Lemma

The statement of proposition VII,30 in Euclid's *Elements* is just that of Euclid's Lemma: 'If two numbers by multiplying one another make some number, and any prime number measure the product, it will also measure one of the original numbers.' The *proof*, however, only uses the assumption that the number measuring the product is prime to deduce that it is *relatively* prime to each of the original factors. It thus establishes the following stronger result, stated earlier in Chapter 4.

Theorem: *If a divides bc and is relatively prime to b, then a divides c.*

Proof (a modern paraphrase of Euclid's argument): Suppose a divides bc, say $bc = ad$, and a is relatively prime to b. Then a must be the least natural number that, when multiplied by d, yields a multiple of c; that is, ad must be the least common multiple of c and d. For let f denote the least such number, and suppose $fd = ec$. By the division algorithm, $a = qf + r$ for some q, r with $0 \leq r < f$, so $bc = ad = qfd + rd = qec + rd$. Hence $rd = bc - qec = (b - qe)c$. By the minimality of f, r must equal 0, so $a = qf$ and (since $c \neq 0$) $b = qe$. q is thus a common factor of a and b, which implies that $q = 1$. Therefore $a = f$, as claimed.

To finish the proof, Euclid appealed to his proposition VII,20 (whose proof, however, was faulty; cf. footnote 1 in Chapter 4): a is the least natural number for which $a/b = c/d$, so a divides c. Alternatively, one may apply the division algorithm again to deduce that $c = pa + s$ for some p, s with $0 \leq s < a$. Then, $s = c - pa$, so

$$sd = cd - pad = cd - pbc = (d - pb)c.$$

That is, sd is a multiple of c, so by the minimality of a, $s = 0$. Thus $c = pa$, so a divides c. q.e.d.

Whichever method is used to complete the proof above, the argument as a whole invokes the division algorithm twice, since (again as noted in footnote 1 of Chapter 4) a correct proof of Euclid's VII,20 employs the division algorithm.

By contrast, the next proof (from Rademacher and Toeplitz 1957, pp. 71–72) of the weaker form of Euclid's Lemma does so only once, to show that the least common multiple of two numbers divides any common multiple of them, but it makes use of an additional fact (displayed as (7) below) not employed in Euclid's argument.

Second proof: If m is the least common multiple of two numbers a and b and M is any common multiple of them, then m divides M; for, by the division algorithm, $M = qm + r$ for some q and r with $0 \leq r < m$, so the minimality of m implies that $r = M - qm$ must be 0. In particular, m must divide ab, say $md = ab$, where

$$d \text{ must divide both } a \text{ and } b. \tag{7}$$

(For if $m = ka = lb$, then $md = kad = ab$ and $md = lbd = ab$, so $kd = b$ and $ld = a$.) Now suppose the prime p divides BC, and let L be the least common multiple of p and B. Since BC and pB are both common multiples of p and B, L divides both of those products — say $LE = BC$ and $LF = pB$. By (7) above, F divides both p and B, and since p is prime, either $F = 1$ or $F = p$; that is, either $L = pB$ or $L = B$. In the former case, $LE = pBE = BC$, so $pE = C$, that is, p divides C. In the latter case, p divides B, since L is a multiple of p. q.e.d.

That the quantity d in (7) is actually the *greatest* common divisor of a and b is nowhere used in the proof above. However, Euclid's Lemma is an almost immediate consequence of the following well-known characterization of greatest common divisors.

Linear representation theorem: *If d is the greatest common divisor of a and b, then there are integers m and n, exactly one of which is positive, for which $d = ma + nb$.*

Proof of Euclid's lemma from the linear representation theorem: If p is a prime that divides bc but not b, then the greatest common divisor of p and b is 1. Therefore $1 = mp + nb$ for some integers m and n, so $c = mpc + nbc$. Since p divides each summand on the right, p divides c. (Exactly the same argument holds if p is not necessarily prime, but merely relatively prime to b.)

The argument just given forms the conclusion of two distinct proofs of Euclid's Lemma, which differ in how the linear representation theorem itself is derived.

Third proof (summarized from Courant and Robbins 1941, pp. 45–47): The representation of the greatest common divisor of integers a and b as an integral linear combination of them is obtained constructively by examining the proof of the Euclidean algorithm (proposition VII,2 in Euclid's *Elements*). That proof, and the implementation of the algorithm to compute m and n explicitly, involves iterated application of the division algorithm, in which *the number of iterations required is not fixed*, as in the two proofs given earlier, but depends on the values of a and b.[2] q.e.d.

Alternatively, the linear representation of the greatest common divisor may be demonstrated non-constructively as follows.

Fourth proof: The set I of *all* linear combinations $ma + nb$, as m and n range over all integers, is an ideal within the ring of integers. Let d be an element of I whose absolute value is minimal. A single application of the division algorithm shows that d must divide every element of I, so in particular it must divide a and b. But *any* common divisor of a and b must also divide every element of I, including d. Therefore d must be a greatest common divisor of a and b (as must $-d$, so d may be taken to be positive without loss of generality). q.e.d.

[2] Of course, the division algorithm itself involves iterated *subtraction*, where the number of iterations likewise depends on the values of the dividend and divisor.

A priori, the greatest common divisor of a and b is just that, the common divisor which is the largest. But one important consequence of the linear representation theorem is the following property of the greatest common divisor.

Divisibility property of the gcd: The greatest common divisor of a and b is divisible by every common divisor of a and b.

Proof. Let c be a common divisor of a and b. Express the greatest common divisor d of a and b as $d = ma + nb$. Then, since c divides both a and b, c divides d as well.

Weintraub (in Weintraub 2008) gave the following proof of Euclid's lemma from the divisibility property of the gcd (which itself may be proved in various ways).

Fifth proof: Consider ac and bc. They have a greatest common divisor d. Now c divides both ac and bc, so by the divisibility property of the gcd, c divides d. Write $d = cz$. Now d divides bc, that is, cz divides bc, so z divides b. Similarly, cz divides ac, so z divides a. But a and b are assumed to be relatively prime, so $z = 1$ and $d = c$. Now a certainly divides ac, and a divides bc by hypothesis, so a divides d by the divisibility property of the gcd again; since $c = d, a$ divides c.

6.2 Indirect proofs of the FTA and Euclid's Lemma

The Fundamental Theorem of Arithmetic may also be proved outright, without first proving Euclid's Lemma, through inductive arguments by *reductio*. Two such proofs, the first by Ernst Zermelo and the second by Gerhard Klappauf,[3] are reproduced in Scholz (1961). Both begin by presuming, contrary to the statement of the FTA, that there are integers with distinct prime factorizations, among which there must be some least integer m. Zermelo then argued as follows.

Sixth proof: Suppose $m = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_s$, with $p_1 \leq p_2 \leq \cdots \leq p_k$ and $q_1 \leq q_2 \leq \cdots \leq q_s$. By the minimality of m, $p_1 \neq q_1$, so without loss of generality we may suppose that $p_1 < q_1$. Then the number

$$n = m - p_1 q_2 \cdots q_s = p_1(p_2 \cdots p_k - q_2 \cdots q_s) = (q_1 - p_1)(q_2 \cdots q_s)$$

is less than m, and so must possess a unique prime factorization. Since p_1 is less than every q_i, it must therefore divide $q_1 - p_1$. But then p_1 would divide q_1, which is prime. Since $1 < p_1 < q_1$, that is impossible. q.e.d.

In the paper in which he presented the proof just given, Zermelo stated that his reason for doing so was to show that even in elementary number theory it was

[3] Published originally in Zermelo (1934) and Klappauf (1935), respectively.

possible to simplify the proofs.[4] His proof, in turn, then stimulated Klappauf to show that the method Zermelo had used to produce the counterexample n could be further simplified.

Seventh proof: Let m be as in Zermelo's proof and consider the remainders r_i, for $i = 1, \ldots, s$ that are obtained when each q_i is divided by p_1. We have $q_i = a_i p_1 + r_i$, where each $r_i < p_1$. Since $p_1 < q_i$ for each i, every a_i must be positive; and since each q_i is a prime different from p_1, every r_i is also positive. Hence $m = q_1 q_2 \cdots q_s$ can be written as $m = Ap_1 + R$, where $R = r_1 r_2 \cdots r_s$ and A and R are both positive. Since p_1 divides m, it must also divide R. But p_1 cannot divide any r_i; so factoring each r_i into primes yields a factorization of R that is distinct from the factorization involving p_1. Since $R < m$ that contradicts the minimality of m. q.e.d.

Unlike Zermelo's proof, Klappauf's employs the division algorithm. Moreover, in Klappauf's proof $r_1 = q_1 - a_1 p_1 \leq q_1 - p_1$, and $r_i < q_i$ for $i \geqslant 2$, so the number R used therein to contradict the minimality of m is less than the number $n = (q_1 - p_1) q_2 \cdots q_s$ used for that purpose in Zermelo's proof.

Euclid's Lemma may also be proved by *reductio*. Indeed, Gauss did so (for the contrapositive statement) in his *Disquisitiones Arithmeticae*. His proof, presented next below, is actually a double *reductio* that invokes the division algorithm thrice.

Eighth proof: Gauss first showed by *reductio* that no prime p can divide a product of two smaller positive integers. For suppose to the contrary that p is a prime that divides such a product, and let $r < p$ be the least positive integer for which there exists a positive integer $s < p$ such that p divides rs. Then $r \neq 1$ (since $s < p$), so r does not divide the prime p. Hence by the division algorithm, $p = qr + t$, where $0 < t < r$. But then $ts = ps - qrs$ is divisible by p, contrary to the minimality of r.

To complete the proof of Euclid's Lemma, suppose then (again by *reductio*) that a prime p divides bc but neither b nor c. Then the division algorithm gives $b = q_1 p + r_1$ and $c = q_2 p + r_2$, with $0 < r_1, r_2 < p$. So bc can be expressed in the form $qp + r_1 r_2$. That is, $r_1 r_2 = qp - bc$, which is a multiple of p if bc is; but that contradicts Gauss's earlier result. q.e.d.

Another *reductio* proof of Euclid's Lemma, in the strong form stated by Euclid, was given by Daniel Davis and Oved Shisha in a little-known article in *Mathematics Magazine* (Davis and Shisha 1981).[5] The last of the proofs to be considered here, it is an elegant exemplar of purity of method.

[4] He noted that he had first communicated his proof around 1912, in correspondence with A. Hurwitz, E. Landau and others, and was stimulated to publish it after reading the proof in the German edition (1933) of Rademacher and Toeplitz's book (the second proof given above), unaware when he did so that a proof similar to his had been published six years earlier by Helmut Hasse (Hasse 1928). In addition, F.A. Lindemann had published another similar, but somewhat more complicated, proof just the year before (Lindemann 1933).

[5] Their paper actually gave five slightly variant proofs.

Ninth proof: Assume that there is a triple (A, B, C) of positive integers for which both of the following properties hold:

$P_1(A, B, C)$. A divides BC but is relatively prime to B.
$P_2(A, B, C)$. A divides BC but does not divide C.

Then for $i = 1, 2$, it follows directly that

$$(1.i) \qquad \text{If } P_i(A, B, C) \text{ and } B > A, \text{ then } P_i(A, B - A, C)$$

and

$$(2.i) \qquad \text{If } P_i(A, B, C), \text{ say } BC = AD, \text{ then } P_i(B, A, D).$$

Proof of (1.1): If A divides BC and $B > A$, then A divides $(B-A)C = BC-AC$; and if A is relatively prime to $B = (B - A) + A$ it must also be relatively prime to $B - A$.

Proof of (1.2): If A divides BC but not C, then A divides $(B-A)C = BC-AC$ but not C.

Proof of (2.1): If $BC = AD$ and A is relatively prime to B, then B divides AD and is relatively prime to A.

Proof of (2.2): If $BC = AD$ but A does not divide C, then B divides AD but does not divide D.

Among all triples (A, B, C) satisfying P_1 and P_2 there is at least one, say (A_1, B_1, C_1), that minimizes $A + B + C$. Then by P_2, $A_1 \neq 1$, so by P_1, $A_1 \neq B_1$. By $(2.i)$, the triple (A_1, B_1, D) also satisfies P_1 and P_2, and $B_1C_1 = A_1D$; so if $B_1 < A_1$, then $D < C_1$ and therefore $A_1 + B_1 + D < A_1 + B_1 + C_1$, contrary to the minimality property of (A_1, B_1, C_1). The only remaining possibility is $B_1 > A_1$. Then by $(1.i)$, the triple (A_1, B_1-A_1, C_1) also satisfies P_1 and P_2; but $B_1-A_1 < B_1$, so $A_1+(B_1-A_1)+C_1 = B_1+C_1 < A_1+B_1+C_1$, again contrary to the minimality property of (A_1, B_1, C_1). Hence by *reductio*, no triple (A, B, C) satisfying both P_1 and P_2 exist. That is, if A divides BC but is relatively prime to B, then A must divide C. q.e.d.

This proof of Davis and Shisha is distinguished above all by its economy of means, for it employs nothing more than subtraction and the concepts involved in the statement of Euclid's Lemma (divisibility and relative primality).

6.3 Summary

It should be clear from the commentary above that the two proofs of the first part of the FTA considered in this chapter, and the nine proofs of the second part, are all structurally distinct. Moreover, they exemplify several of the rationales for presenting alternative proofs enumerated in Chapter 2: the desires to simplify, to minimize conceptual prerequisites, to extend to broader contexts, to achieve methodological purity, and to find new routes to a goal. As to proofs of the first part, the second proof is direct while the first proof is a proof by *reductio*. As to the second part, Gauss's proof extended that of Euclid; the second, sixth and seventh proofs, and especially the ninth, exhibit various forms of simplification; and the third and fourth proofs introduce a concept (that of representing the greatest common divisor of two integers as a linear combination of them) that is foreign to all the others, while the fifth proof deliberately avoids using that concept.

It is enlightening to examine these proofs in the context of generalizations to commutative ring theory. As to the proofs of the first part, the first proof leads directly to the concept of a Noetherian ring, and directly generalizes to show that every element in a Noetherian integral domain has a (that is, at least one) factorization into primes. The second proof, while simpler, is one that is restricted to the positive integers.

As to the proofs of the second part, again the most direct proofs, the sixth, seventh, and ninth, are restricted to the positive integers.

The other proofs generalize to commutative rings of various kinds. By definition, a Euclidean domain is one in which there is a division algorithm, properly interpreted, and these proofs show that Euclid's lemma holds in Euclidean domains, and hence that these are unique factorization domains (that is, that the analog of the FTA holds in them). By definition, a principal ideal domain is one in which an appropriate generalization of the linear representation theorem holds, and so the third and fourth proofs, which rely on that concept, show that every principal ideal domain is a unique factorization domain. Since there are principal ideal domains that are not Euclidean, this provides a further generalization. Furthermore, not every unique factorization domain is a principal ideal domain, so the fifth proof generalizes still further (though in this case one needs some other argument to show that the divisibility property of the gcd holds). The second proof, which introduces the concept of the least common multiple, is stated for the positive integers, and directly generalizes to Euclidean domains. But that same concept is fruitful in the more general contexts we have just described, and that proof can be modified to be valid in them as well.

References

Collison, M.J.: The unique factorization theorem, from Euclid to Gauss. Math. Mag. **53**(2), 96–100 (1980)

Courant, R., Robbins, H.: What is Mathematics? Oxford U., Oxford et al. (1941)

Davis, D., Shisha, O.: Simple proofs of the fundamental theorem of arithmetic. Math. Mag. **54**, 18 (1981)

Gauss, C.F.: Disquisitiones Arithmeticae (1801). In his Werke, I, 1–466. W.F. Kaestner, Göttingen (1863)

Hasse, H.: Über eindeutige Zerlegung in Primelemente oder in Primhauptideale in Integritäts Bereichung. J. reine. u. angew. Math. **159**, 3–12 (1928)

Klappauf, G.: Beweis des Fundamentalsatz der Zahlentheorie. Jahresb. DMV **45**, 130 (1935)

Lindemann, F.A.: The unique factorization of a positive integer. Quarterly J. Math. **4**, 319–320 (1933)

Rademacher, H., Toeplitz, O.: The Enjoyment of Mathematics. Princeton U.P., Princeton (1957)

Scholz, A.: Einführung in die Zahlentheorie. De Gruyter, Berlin (1961)

Weintraub, S.H.: Factorization: Unique and Otherwise. AK Peters, New York (2008)

Zermelo, E.: Elementare Betrachtungen zur Theorie der Primzahlen. Nachr. Gesell. Wissen. Göttingen (Neue Folge) **1**, 43–46 (1934)