S

T

U

V

W

Z

Compact Textbooks in Mathematics

Hardy Grant
Israel Kleiner

Turning Points in the History of Mathematics

Birkhäuser

Compact Textbooks in Mathematics

For further volumes:
http://www.springer.com/series/11225

Birkhäuser

H

I

K

L

M

Compact Textbooks in Mathematics

Hardy Grant
Israel Kleiner

Turning Points in the History of Mathematics

Birkhäuser

Index

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics*, Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1, © Springer Science+Business Media, LLC 2015

Hardy Grant
Department of Mathematics and Statistics
York University
Toronto, Ontario
Canada

Israel Kleiner
Department of Mathematics and Statistics
York University
Toronto, Ontario
Canada

ISSN 2296-4568
ISSN 2296-455X (electronic)
ISBN 978-1-4939-3263-4
ISBN 978-1-4939-3264-1 (eBook)
DOI 10.10.1007/978-1-4939-3264-1
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2015952852

Birkhäuser

Printed on acid-free paper
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

The development of mathematics has not followed a smooth or continuous curve, although in hindsight we may think so. As the mathematician and historian of mathematics Eric Temple Bell (1883–1960) said: “Nothing is easier … than to fit a deceptively smooth curve to the discontinuities of mathematical invention” [1, p. viii]. In fact, there have been dramatic insights and breakthroughs in mathematics throughout its history, as well as what seemed for a time to be insurmountable stumbling blocks—both leading to major shifts in the subject. And then—for the most part—there have been relatively “routine” developments, from whose importance we do not wish to detract.

Here are two “nonroutine” examples:

a. The invention (discovery?) of noneuclidean geometry—a breakthrough which was about two millennia in the making (ca 300 BC—ca 1830), and which culminated in the resolution of “the problem of the fifth postulate.” This brought about a reevaluation of the nature of geometry and its relationship to the physical world and to philosophy, as well as a reconsideration of the nature of axiomatic systems. See ▶ Chapter 7.
b. The introduction, around the mid-eighteenth century, of “foreign objects”, such as irrational and complex numbers, into number theory, to be followed in the late nineteenth century by the founding of a new subject—algebraic number theory. These developments paved the way for splendid achievements of modern mathematics, including, to take a familiar example, the resolution of the problem, stated in the 1630s, concerning the unsolvability in integers of Fermat’s equation $x^n + y^n = z^n$, $n > 2$. The proof of unsolvability, given by Andrew Wiles in 1994, required most of the grand ideas which number theory had evolved during the twentieth century. See ▶ Chapter 6 and [4].

We aim in this book to discuss some of these major turning points—transitions, shifts, breakthroughs, discontinuities, revolutions (if you will)—in the history of mathematics, ranging from ancient Greece to the present [2, 3]. Among those which we consider are the rise of the axiomatic method (▶ Chapter 1), the wedding of algebra and geometry (▶ Chapter 4), the taming of the infinitely small and the infinitely large (▶ Chapter 5), the passage from algebra to algebras (▶ Chapter 8), and the revolutions resulting in the late nineteenth and early twentieth centuries from Cantor’s creation of transfinite set theory (▶ Chapters 9 and 10). The historical origin of each turning point is discussed, as well as some of the resulting mathematics.

The above examples, and others discussed in this book, highlight the great drama inherent in the evolution of mathematics. Teachers of this grand subject will benefit from reflecting on this important aspect of it, focusing on the big ideas in its development—though not, of course, to the neglect of “routine” mathematics. They should pass on to students—at some point in their studies—at least the spirit, if not always the content, of these ideas. In particular, students should be made aware that not every fact, technique, idea, or theory is as important, and should receive as much emphasis, as every other. If this thought is not conveyed to them, our teaching will do justice neither to the students nor to the subject.

39. Laugwitz, D.: Bernhard Riemann 1826–1866: Turning Points in the Conception of Mathematics. Birkhäuser, Boston (1999). (Translated from the German by A. Shenitzer.)
40. Manin, Yu.I.: How convincing is a proof? Math. Intell. **2**(1), 17–18 (1979)
41. Mazur, I.: Enlightening Symbols: A Sort History of Mathematical Notation and Its Hidden Powers. Princeton University Press, Princeton (2014)
42. Norwood, F.H.: Long proofs. Am. Math. Mon. **89**, 110–112 (1982)
43. Spencer, I.: Short theorems with long proofs. Am. Math. Mon. **90**, 365–366 (1983)
44. Stahl, S.: Real Analysis: A Historical Approach. Wiley, New York (1999)
45. Swart, E.R.: The philosophical implications of the four-color problem. Am. Math. Mon. **87**, 697–707 (1980)
46. Tymoczko, T.: The four-color problem and its philosophical significance. J. Phil. **76**(2), 57–83 (1979)
47. Tymoczko, T.: Computers, proofs and mathematicians: A philosophical investigation of the four-color proof. Math. Mag. **53**, 131–138 (1980)

3. Borwein, J., Bailey, D.: Mathematics by Experiment: Plausible Reasoning in the 21st Century, 2nd edn. A K Peters, Natick (2008)
4. Bottazzini, U.: The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass. Springer, New York (1986)
5. Cayley, A.: Chapters in the Analytic Geometry of (n) Dimensions. Camb. Math. J. **IV**, 119–127 (1845)
6. Chong, T., Leong, Y.K.: An interview with Jean-Pierre Serre. Math. Intell. **8**(4), 8–13 (1986)
7. Crilly, T., Cayley, A.: Mathematician Laureate of the Victorian Age. Johns Hopkins University Press, Baltimore (2005)
8. De Millo, R.A., Lipton, R.J., Perlis, A.J.: Social processes and proofs of theorems and programs. Comm. ACM **22**, 271–280 (1979)
9. Edwards, C.H.: The Historical Development of the Calculus. Springer, New York (1979)
10. Glaisher, J.W.: On the early history of the signs "+" and "−", and on the early German arithmeticians. Messenger Math. **51**, 1–148 (1921–1922)
11. Hankins, T.L.: Sir William Rowan Hamilton. Johns Hopkins University Press, Baltimore (1980)
12. Hawkins, T.: Lebesgue's Theory of Integration: Its Origins and Development. Chelsea Publishing Co., New York (1970)
13. Katz, V.: A History of Mathematics: An Introduction, 3rd edn. (2009)
14. Kline, M.: Mathematical Thought from Ancient to Modern Times. Oxford University Press, Addison-Wesley, Boston (1972)
15. Van der Waerden, B.L.: Hamilton's discovery of quaternions. Math. Mag. **49**, 227–234 (1976)

Further Reading

16. Appel, K., Haken, W.: The four-color problem. In: Steen L.A. (ed.) Mathematics Today—Twelve Informal Essays, pp. 153–190. Springer, New York (1978)
17. Boyer, C.B.: Fractional indices, exponents, and powers. National Math. Mag. **18**, 81–86 (1943)
18. Bressoud, D.: A Radical Approach to Real Analysis. Mathematical Association of America, Washington DC (1994)
19. Cajori, F.: The spread of Newtonian and Leibnizian notations of the calculus. Bull. Am. Math. Soc. **27**, 453–458 (1921)
20. Cajori, F.: The history of notations of the calculus. Ann. Math. **25**, 1–46 (1924)
21. Cajori, F.: Leibniz: The master builder of notations. Isis. **7**, 412–429 (1925)
22. Cajori, F.: A History of Mathematical Notations, 2 vols. Open Court, La Salle (1928 & 1929)
23. Crowe, M.J.: A History of Vector Analysis. University of Notre Dame Press, Notre Dame (1967)
24. Davis, P.J., Hersh, R.: The Mathematical Experience. Birkhäuser, Boston (1981). (Revised Study Edition, with E. Marchisotto, published in 1995.)
25. Detlefsen, M., Luker, M.: The four-color theorem and mathematical proof. J. Phil. **77**, 803–820 (1980)
26. Epstein, D., Levy, S.: Experimentation and proof in mathematics. Notices Am. Math. Soc. **42**, 670–674 (2005)
27. Gelbaum, B.R., Olmsted, J.M.H.: Counterexamples in Analysis. Holden-Day, San Francisco (1964)
28. Grabiner, J.V.: Is mathematical truth time-dependent? Am. Math. Mon. **81**, 354–365 (1974)
29. Grattan-Guinness, I. (ed.): From the Calculus to Set Theory. 1630–1910, Princeton University Press, Princeton (2000)
30. Hairer, E., Wanner, G.: Analysis by its History. Springer, New York (1996)
31. Haken, W.: An attempt to understand the four-color problem. J. Graph Theory **1**, 193–206 (1977)
32. Hersh, R.: What is Mathematics, Really? Oxford University Press, New York (1997)
33. Katz, V., Parshall, K.: Taming the Unknown: A History of Algebra from Antiquity to the Early Twentieth Century. Princeton University Press, Princeton (2014)
34. Klymchuk, S.: Counterexamples in Calculus. Mathematical Association of America, Washington DC (2010)
35. Kolata, G.: Mathematical proofs: The genesis of reasonable doubt. Science. **192**, 989–990 (1976)
36. Lakatos, I.: Proofs and Refutations. Cambridge University Press, Cambridge (1976)
37. Lakatos, I.: A renaissance of empiricism in the recent philosophy of mathematics? In: Tymoczko T. (ed.) New Directions in the Philosophy of Mathematics, pp. 29–48. Birkhäuser, Boston (1986)
38. Lakatos, I.: What does a mathematical proof prove? In: Tymoczko T. (ed.) New Directions in the Philosophy of Mathematics, pp. 153–162. Birkhäuser (1986)

The book contains ten chapters, more or less of equal length, though not of equal difficulty. They describe only a small number of "turning points" in the history of mathematics, and we have appended an 11th chapter which suggests "Some Further Turning Points" to pursue. Each chapter contains about ten "problems and projects", most of which are intended to deepen or extend the material in the text. At the end of each chapter there is a substantial list of references, whose aim is to elaborate, enhance, and exemplify the material in the text proper. Finally, the book has a comprehensive index.

This book can be read by a person with some mathematical background who is interested in getting a nontraditional look at aspects of the history of mathematics. It can also be used in history-of-mathematics courses, especially those centered around the important idea of "turning points." Moreover, since appreciation of the historical development of the central ideas of mathematics enhances, we strongly believe, one's understanding and appreciation of the subject, this book can serve as a text in a capstone course for mathematics majors, a course that will integrate and "humanize" at least some of their knowledge of mathematics by placing it in historical perspective. In any such course our book will probably need to be supplemented by additional technical material; a teacher will know best when and how to use this "extra" material in his or her particular classroom setting. Teachers are resourceful and will likely use the book in ways we have not anticipated.

One of the reviewers of our book said the following: "I see the value of the manuscript in its role as a 'starter' to ignite love for the history of maths and to give a first overview. It is a good 'teaser'." We hope that readers' experiences will justify this assessment.

We want to thank Chris Tominich, Assistant Editor, Birkhäuser Science, for his outstanding cooperation in seeing this book to completion, and Ben Levitt, Birkhäuser Science Editor, for his cordial and efficient support.

I (HG) want to thank three generations of my family—my dear sisters Nancy and Kathy, and my cherished nephews, nieces-in-law, and great-nieces Ross, Sevaun, Zada, and Alyn, and Ian, Lynn, and Charlotte—for love, inspiration, and many good times.

And I (IK) want to thank my dear wife of 50 years, Nava, for her support and encouragement over these many years. I have also gained invaluable perspective in seeing our children and grandchildren—Ronen, Leeor, Tania, Ayelet, Howard, Tamir, Tia, Jordana, Jake, and Elise—grow, mature, and thrive.

References

1. Bell, E.T.: The Development of Mathematics, 2nd ed. McGraw-Hill, New York (1945)
2. Gillies, D. (ed.): Revolutions in Mathematics. Clarendon Press, Oxford (1992)
3. Laugwitz, D.: Bernhard Riemann 1826–1866: Turning Points in the Conception of Mathematics. Birkhauser, New York (2008)
4. Singh, S.: Fermat's Enigma: The Quest to Solve the World's Greatest Mathematical Problem. Penguin, New York (1997)

Hardy Grant and Israel Kleiner

Contents

Jonathan Borwein (1951–)

The methods of this field are thus for the most part akin to those of the scientist: experimenting, much of it via the computer and its increasingly sophisticated tools, formulating hypotheses, and testing these by further experimentation. These various activities—short of proof—are publishable, following the usual reviewing process. Not that proof is to be abandoned, but the focus is elsewhere. As Borwein, who calls himself a "computer-assisted fallibilist", asserts [2, p. 35]:

> In my view, it is now both necessary and possible to admit quasi-empirical inductive methods fully into mathematical argument. In doing so we will enrich mathematics…. Mathematics is primarily about *secure knowledge,* not proof …. Proofs are often out of reach— but understanding, even certainty, is not.

As an illustration Borwein gives the following example [2, p. 37]:

> Given an interesting identity buried in a long and complicated paper on an unfamiliar subject, which would give you more confidence in its correctness: staring at the proof, or confirming computationally that it is correct to 10,000 decimal places? Here is such a formula [which arose in quantum field theory]:

$$\left[\frac{24}{7\sqrt{7}}\right]\int_{\frac{\pi}{3}}^{\frac{\pi}{2}}\log\left|\left(\tan t+\sqrt{7}\right)\Big/\left(\tan t-\sqrt{7}\right)\right|dt =$$
$$\sum_{n=0}^{n=\infty}\left[\frac{1}{(7n+1)^2}+\frac{1}{(7n+2)^2}+\frac{1}{(7n+3)^2}+\frac{1}{(7n+4)^2}+\frac{1}{(7n+5)^2}+\frac{1}{(7n+6)^2}\right]$$

Does this new subject represent a paradigm shift in mathematics?

References

1. Bailey, D., Borwein, J., et al.: Experimental Mathematics in Action. A K Peters, Natick (2007)
2. Borwein, J.: Implications of experimental mathematics for the philosophy of mathematics. In: Gold, B., Simons, A. (eds.) Proof & Other Dilemmas: Mathematics and Philosophy, pp. 33–59. Mathematical Association of America (2008)

Similar objections were raised to very long *traditional* proofs of theorems, for example, the proof (in the 1960s) of the Feit-Thompson theorem describing the solvability of finite groups of odd order. Speaking of this theorem, and other results whose proofs are very long, the famed mathematician Jean-Pierre Serre observed [6, p. 11]:

> What shall one do with such theorems, if one has to use them? Accept them on faith? Probably. But it is not a very comfortable situation.

Largely as a result of these developments, a novel philosophy of mathematical proof, called "quasi-empiricist proof" or "proof as a social process", has emerged. Its essence, according to its advocates, is that *proofs are not infallible.* Thus mathematical theorems cannot be guaranteed absolute certainty. And this applies not only to the theorems requiring very long proofs or assistance of a computer but also to many "run of the mill" cases [8].

11.5 Experimental Mathematics: From Humans to Machines

> One of the greatest ironies of the information technology revolution is that while the computer was conceived and born in the field of pure mathematics, through the genius of giants such as John von Neumann and Alan Turing, until recently this marvelous technology had only a minor impact within the field that gave it birth [3, p. 1].

These words were published in 2008. But during the last two decades or so of the twentieth century, the computer did in fact make wide and deep inroads into mathematics. It invigorated old fields, and it stimulated, or was instrumental in, the founding of new ones. In the latter category was "experimental mathematics", established around 1990 by David Bailey (1948–), Jonathan Borwein (1951–), and others. This was a major departure from the traditional understanding of the mathematical enterprise, for it entailed "the utilization of advanced computer technology in mathematical research" [1, p. 2].

In 1991 a new journal, *Experimental Mathematics,* was launched. Its founders noted that it differs from the traditional mathematics journals in that its focus is "not only [on] theorems and proofs but also [on] the way in which they have been or can be reached. … The objective of *Experimental Mathematics* is to play a role in the discovery of formal proofs, not to displace them" [2, p. 671]. Bailey and Borwein give a more detailed "definition" of experimental mathematics [1, pp. 2, 3]:

> By experimental mathematics we mean the methodology of doing mathematics that includes the use of computations for [1, p. 3]:
> 1. Gaining insight and intuition.
> 2. Discovering new patterns and relationships.
> 3. Using graphical displays to suggest underlying mathematical principles.
> 4. Testing and especially falsifying conjectures.
> 5. Exploring a possible result to see if it is worth a formal proof.
> 6. Suggesting approaches for formal proof.
> 7. Replacing lengthy hand derivations with computer-based derivations.
> 8. Confirming analytically based results.

Kenneth Appel (1932–2013)

But not everyone was pleased with these developments, these "exceptions and irregularities". Some called such functions "pathological", others gave them less pleasant designations. Thus Charles Hermite asserted (in 1893) that he "turn[ed] away with fright and horror from this lamentable evil of functions … " [14, p. 973]. Henri Poincaré was more specific (1899) [14, p. 973]:

> Logic sometimes makes monsters. For half a century we have seen a mass of bizarre functions which appear to be forced to resemble as little as possible honest functions which serve some purpose. More of continuity, or less of continuity, more derivatives, and so forth …. In former times when one invented a new function it was for a practical purpose; today one invents them purposely to show up defects in the reasoning of our fathers and one will deduce from them only that.

Hermite and Poincaré did not prevail, of course. The work of Riemann, Weierstrass, and others (in the second half of the nineteenth century) in analysis necessitated—once again—a reexamination of its foundations, leading to the "arithmetization of analysis" [4, 14].

11.4 The Nature of Proof: From Axiom-Based to Computer-Assisted

The later decades of the twentieth century saw the solution of major outstanding mathematical problems—including the Kepler conjecture, the four-color conjecture, the Bieberbach conjecture, Fermat's Last "Theorem", the Feit-Thompson conjecture, the problem of classification of all finite simple groups, and the Poincaré conjecture (this last was solved in 2006). The computer played a major role in establishing some of these conjectures—and several others. This has occasioned a rethinking of the meaning and role of proof in mathematics.

The catalyst was the computer-aided proof (1976) of the four-color theorem by Kenneth Appel and Wolfgang Haken. The proof required the verification, by computer, of 1482 distinct configurations. Some critics argued that this proof (and others like it) was a major departure from tradition. They advanced several reasons:

a. The proof contained thousands of pages of computer programs *that were not published* and thus were not open to the traditional procedures of verification by the mathematical community. In particular, how can a referee check the entire proof?
b. Both computer hardware and computer software are subject to error. Is the computer, then, an experimental tool?

Karl Weierstrass (1815–1897s)

Bernhard Riemann (1826–1866)

adequate for the mathematics of the nineteenth century. For instance, in 1829 Peter Lejeune Dirichlet gave an example of a function—namely, $D(x) = c$, if x is rational, and $D(x) = d$, if x is irrational (c and d unequal real numbers)—which is not an analytic expression. Moreover D(x) is discontinuous everywhere. Eighteenth-century mathematicians believed—and some "proved"!—that every function is continuous, except perhaps at finitely many points. But in the mid-nineteenth century Bernhard Riemann and Karl Weierstrass gave examples of functions that are everywhere continuous but nowhere differentiable, and Riemann exhibited a highly discontinuous Riemann-integrable function. These examples were not given for idle play; they were important in connection with the study of Fourier-series representation of functions [4, 9].

The character of analysis was evolving. Since the seventeenth century its processes had been assumed to apply to "all" functions, but it now turned out that they are restricted to particular *classes* of functions. In fact, the investigation of various classes of functions—such as continuous functions, semi-continuous functions, differentiable functions, functions with nonintegrable derivatives, integrable functions, monotonic functions, continuous functions that are not piecewise monotonic—became a principal concern of analysis. Whereas mathematicians had formerly looked for order and regularity in analysis (calculus), some now took delight in discovering exceptions and irregularities (See [12, p. 119]).

Axiomatics—Euclid's and Hilbert's: From Material to Formal

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics*, Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_1, © Springer Science+Business Media, LLC 2015

1.1 Euclid's *Elements*

The axiomatic method is, without doubt, the single most important contribution of ancient Greece to mathematics. The explicit recognition that mathematics deals with abstractions, and that proof by deductive reasoning from explicitly stated postulates offers a foundation for mathematics, was indeed an extraordinary development. When, how, and why this came about is open to conjecture. Various reasons—both internal and external to mathematics (Raymond Wilder calls them "hereditary" and "environmental" stresses, respectively [14])—have been advanced, with various degrees of certainty, for the emergence of the axiomatic method in ancient Greece. Among the suggested causes are:

a. The nature of Greek society. Of course people have always—if often unconsciously—used "axioms" in conversations—shared presumptions from which one participant urges conclusions on the others. Two characteristic features of Greek experience, both dating from the fifth century BC, spurred reflection on these everyday occurrences. The courts of law and the citizen assemblies created by the (limited) democracy conferred high value on the techniques and strategies of skilful persuasion. A science of "rhetoric" developed, in which argument proceeded in specific modes and stages, analogous to the successive steps in a euclidean proof (see [10], ► Chapter 4).
b. The predisposition of the Greeks to the kind of philosophical inquiry in which answers to ultimate questions are of prime concern. For example, the attempt by Parmenides (ca. 515–450 BC), the founder of the "Eleatic" school of philosophy, to show that all of ultimate reality is an unchanging unity is generally taken by modern scholars to be the oldest deductive argument that has come down to us [10, p. 102]. Parmenides used the *indirect* method of proof, assuming the denial of his intended conclusion and reaching an untenable outcome. The famous paradoxes of Parmenides' pupil Zeno, which claim to prove that motion is impossible, are (of course) also deductive arguments. Zeno also used the indirect method of proof (see [11], but also [9], in which an alternative thesis is proposed). Both of these thinkers may have consciously worked from explicit assumptions—in effect, axioms—though no hint of these survives. In this connection it is interesting to note the view of the eighteenth-century mathematician and scientist A.-C. Clairaut regarding Euclid's proofs of obvious propositions [8, pp. 618 – 619]:

» It is not surprising that Euclid goes to the trouble of demonstrating that two circles which cut one another do not have a common centre, that the sum of the sides of a triangle which is enclosed within another is smaller than the sum of the sides of the enclosing triangle. This geometer had to convince obstinate sophists who glory in rejecting the most evident truths; so that geometry must, like logic, rely on formal reasoning in order to rebut the quibblers.

c. The desire to decide among contradictory results bequeathed to the Greeks by earlier civilizations [12, p. 89]. For example, the Babylonians used the formula $3r^2$ for the area of a circle, the Egyptians $\left[\frac{8}{9}\times 2r\right]^2$. (There is evidence that the Babylonians also used $3\frac{1}{8}$ as an estimate for π [8, p. 11].) This encouraged the notion of mathematical demonstration, which in time evolved into the deductive method.
d. The need to resolve the "crisis" engendered in the fifth century BC by proof of the incommensurability of the diagonal and side of the square [3]. A fundamental tenet of the Pythagoreans was that all phenomena can be described by numbers, which to them meant positive integers. They developed important parts of geometry with the aid of this principle. In particular, the principle implied that any two line segments a and b are commensurable (have a common measure), that is, that there exists a line segment t such that $a = mt$, $b = nt$, with m and n positive integers. But around 430 BC they proved that the side and diagonal of a unit square are *not* commensurable. (In a modern formulation we would say that $\sqrt{2}$ is irrational.) [7]. This must have been a great shock to them, as it went counter to their philosophy and their mathematics. And it might have provided an important impetus for a critical reevaluation of the logical foundations of mathematics [3(a)].
e. The need to teach. This may have forced the Greek mathematicians to consider the basic principles underlying their subject. It is noteworthy that the pedagogical motive in the formal organization of mathematics is also present in the works of later mathematicians, notably Lagrange, Cauchy, Weierstrass, and Dedekind [8].

Euclid's great merit was to have collected, and arranged brilliantly in a grand axiomatic edifice called *Elements*, much of the mathematics of the previous three centuries (with notable exceptions, such as conic sections). His opus comprises over 450 propositions (theorems), deduced from *five* (!) postulates (axioms), and arranged in thirteen "Books" (chapters). The postulates are:

1. A straight line may be drawn between any two points.
2. A straight line segment may be produced indefinitely.
3. A circle may be drawn with any given point as centre and any given radius.
4. All right angles are equal.
5. If a straight line intersects two other straight lines lying in a plane, and if the sum of the interior angles thus obtained on one side of the intersecting line is less than two right angles, then the straight lines will eventually meet, if extended sufficiently, on the side on which the sum of the angles is less than two right angles.

For over two thousand years, to teach elementary geometry meant to teach it essentially as Euclid had presented it. His masterpiece first appeared *in print* in 1482 (the printing press originated in ca. 1450). More than a thousand editions have appeared since, a profusion superseded

Hermann Grassmann (1809–1877)

A pioneering work on the abstract notion of a vector space of arbitrary dimension was Hermann Grassmann's *Doctrine of Linear Extension* (1844). But this book attracted little attention: it was "philosophical", and it was too abstract for its time. An 1862 edition was better received. An abstract treatment of basic elements of vector space theory was given in 1888 by Giuseppe Peano in his *Geometric Calculus*. See [13, 14].

11.3 Pathological Functions: From Calculus to Analysis

The aim of this section is to indicate some high points in the transition—in the nineteenth century—from calculus to analysis, in which "pathological functions" played a central role.

The calculus invented by Newton and Leibniz (see ▶ Section 11.1, above, and ▶ Chapter 5) was based on variables related by equations, with the focus on geometry: finding areas, volumes, tangents.

The concept of *function* was introduced in the first half of the eighteenth century, and was made central around 1750 by Leonhard Euler, who declared—and showed in his books—that calculus is the branch of mathematics dealing with functions. Euler considered a function to be an (algebraic) "formula"—a so-called "analytic expression". Neither "formula" nor "analytic expression" was defined, but many examples were given. The essential point is that the concept of variable, applied to geometric objects in the seventeenth century, was gradually replaced in the eighteenth by that of function, understood to be an algebraic formula.

The nineteenth century ushered in a period of rigor in various areas of mathematics. Augustin-Louis Cauchy (1789–1857) found the lack of rigor in calculus unsatisfactory, and his textbooks of the 1820s aimed at a remedy. He selected a few fundamental notions (limit, continuity, convergence, derivative, and integral), established the limit concept as the one on which to base all the others, and derived by fairly rigorous means the subject's major results. It is important to note that most of these concepts, as *we* understand them, were either not recognized or not clearly formulated before Cauchy's time. A "new" subject, "analysis", also known as the "theory of functions", was thus born [4].

But there were shortcomings in Cauchy's program. In particular, he conceived a function in the eighteenth-century way—as an analytic expression (a formula). But this was no longer

François Viète (1540–1603)

integers, decimal fractions, exponents, and arithmetic signs—for example, signs for addition and multiplication.

11.2 Space Dimensions: From 3 to n (n >3)

For about two millennia, beginning with ancient Greece, mathematical concepts, results, and theories were often deemed to represent the physical world. In particular, euclidean 3-space was an idealization of the space of our daily experience, and since the latter is 3-dimensional, so was the former. Because space of dimension greater than 3 made no physical sense, it made no mathematical sense. For example, Aristotle stated that "no magnitude can transcend three because there are no more than three dimensions" [14, p. 1028], and John Wallis considered a space of dimension greater than 3 "a monster in nature … less possible than a chimera" [14, p. 1028].

In the nineteenth century mathematics increasingly came to be severed from its intimate relation to the physical world (there were a few earlier examples, for instance complex numbers). Mathematicians began to introduce concepts and derive results with little thought of their application in nature. "The essence of mathematics lies in its freedom" , asserted Georg Cantor, giving expression to this view [14, p. 1031].

In the 1840s William Rowan Hamilton and Arthur Cayley, among others, introduced 4-dimensional geometry [5, 7, 11]. Hamilton, for example, tried to extend the multiplication of complex numbers (pairs of reals) to *triples* of reals – whereupon, he claimed, "there dawned on me the notion that we must admit, in some sense, a *fourth dimension* of space for the purpose of calculating with triples" [15, p. 230], thereby giving birth to the quaternions. Cayley (and others) extended these to the octonions, 8-tuples of reals (see ▶ Chapter 8). Cayley also wrote a paper in 1843 entitled *Chapters in the Analytic Geometry of (n) Dimensions* [5]. We see in these works both algebraic and geometric motivations.

Euclid (fl. ca. 300 BC)

only by the Bible. The *Elements* also inspired Newton to present his masterpiece of physics and cosmology, the *Principia,* axiomatically, and it inspired Spinoza to write his philosophical *chef d'oeuvre,* the *Ethics,* in the same style.

1.2 Hilbert's *Foundations of Geometry*

But despite this influence of the *Elements,* the *practice* of mathematics in the euclidean manner is a rather rare phenomenon in the five thousand-year history of mathematics. It was consciously undertaken for around two hundred years in ancient Greece and was resumed in the nineteenth century. Both of these periods were preceded by centuries of mathematical activity that was often vigorous but rarely rigorous. But for over two millennia there was only one geometry—Euclid's. Its axioms, with the exception of the fifth, were taken to describe an idealization of physical space and were therefore viewed as "self-evident truths", not open to critique. Its theorems, which were logical consequences of the axioms, were therefore also viewed as truths (see ▶ Chapter 7).

The nineteenth century brought a revolution in geometry, both in scope and in depth. New varieties emerged: projective geometry (Girard Desargues' work in 1639 on the subject came to light only in 1845), hyperbolic geometry, elliptic geometry, Riemannian geometry, differential geometry, and algebraic geometry. Jean Victor Poncelet founded synthetic projective geometry in the early 1820s as an independent subject, but lamented its lack of general principles, and the validity of his "principle of duality"—that the truth of theorems is preserved by interchange of "point" and "line"—was questioned. The *consistency* of noneuclidean (hyperbolic) geometry and the relationship of axioms to the physical world were also in debate. And the relative merits of geometric *methods* were contested: the metric versus the projective, the synthetic versus the analytic.

Important new ideas entered geometry: points and lines at infinity, use of complex numbers (cf. complex projective space), use of calculus, extension of geometry to n dimensions, Hermann Grassmann's "calculus of extension", invariants—for example, the "invariant theory of forms" developed by Cayley and Sylvester, and groups—for example, groups of the regular solids. An important development was Felix Klein's proof (1871) that euclidean, hyperbolic, and elliptic geometries are subgeometries of projective geometry. For a time it was said that projective geometry was *all* of geometry [5, p. 239]. This period of profound changes left many mathematicians uneasy. The historian of mathematics Jeremy Gray (1947–) claimed that "signs of anxiety about the nature of geometry run like fissures through late 19th-century mathematics" [5, p. 247].

Euclidean geometry did not escape scrutiny. Although Euclid was the paragon of rigor for more than two thousand years, logical shortcomings were now recognized in his masterpiece *Elements*. For example, his very first proposition in Book I, which presents the construction of an equilateral triangle, has a faulty proof: while Euclid assumed implicitly that two circles, each of which passes through the center of the other, intersect, this observation requires an axiom of continuity, supplied two millennia later by David Hilbert. Gauss pointed out that such concepts as "between", used freely and intuitively by Euclid, must be given an axiomatic formulation.

These challenges were taken up during the last two decades of the nineteenth century by a number of mathematicians. They provided, for projective, euclidean, and noneuclidean geometries, axioms free of the types of blemishes that appear in Euclid's presentation. The first to do this was Moritz Pasch, who wrote an extensive work in 1882 on the foundations of projective geometry. Pasch set out clearly a crucial aspect of modern axiomatics, which departs radically from Euclid's procedure [8, p. 1008]**:**

> » If geometry is to become a genuine deductive science, it is essential that the way in which inferences are made should be altogether independent of the *meaning* of the geometrical concepts, and also of the diagrams; all that need be considered are the relationships between the geometrical concepts asserted by the propositions and definitions.

The most influential work in this genre was Hilbert's *Foundations of Geometry* of 1899. His aim was "to present a *complete* and *simplest possible* system [Hilbert's italics] of axioms [for euclidean geometry], and to derive from these the most important geometrical theorems" [1, p. 344]. To avoid the pitfalls in Euclid's *Elements*—reliance on intuitive arguments, often based on diagrams—Hilbert required *twenty* postulates; Euclid, recall, had *five*. Hilbert listed his axioms under five headings: I. axioms of connection, II. axioms of order, III. axiom of parallels (Euclid's fifth axiom), IV. axioms of congruence, and V. axiom of continuity (Archimedes' axiom).

Crucial was the use, as urged by Pasch, of *undefined* terms, so-called primitive terms. Why are they needed? Because just as one cannot *prove* everything, hence the need for axioms, so one cannot *define* everything, hence the need for undefined terms. They are not uniquely determined; among Hilbert's choices are "point", "(straight) line", and "plane". Euclid defined all three terms, for example, a "point" as "that which has no part"—which is not very informative.

Euclid considered his axioms to be self-evident truths, but Hilbert's are neither self-evident nor true. They are simply the starting points, the basic building blocks, of the theory—assumptions about the relations among the primitive terms of the axiomatic system. The primitive terms are said to be "implicitly" defined by the axioms. As early as 1891 Hilbert highlighted the observation about the arbitrary nature of the primitive terms in the now classic remark that "It must be possible to replace in all geometric statements the words point, line, plane by table, chair, mug" [13, p. 14]. It follows that the axioms, hence also the theorems, are *devoid of meaning.* It is therefore not inappropriate to call Euclid's system "material axiomatics" and Hilbert's system "formal axiomatics" [3(a), p. 63 and 3(b), p. 171].

Hilbert's *Foundations of Geometry* went through ten editions (in the original German), seven in Hilbert's lifetime. It served as a model of what an axiom system should be like, and more broadly, it "demonstrated brilliantly the vitality of the new axiomatic approach to geometry" [1, p. 361]. Garrett Birkhoff and Mary Katherine Bennett wrote (1987) of the *Foundations* that

Some Further Turning Points

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics,* Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_11, © Springer Science+Business Media, LLC 2015

We have discussed in this booklet a number of turning points in the evolution of mathematics, but of course we have not exhausted them all. In this final chapter we suggest five other topics for the reader to pursue—with relatively brief outlines and references:

a. Notation: From Rhetorical to Symbolic
b. Space Dimensions: From 3 to n (n >3)
c. Pathological Functions: From Calculus to Analysis
d. The Nature of Proof: From Axiom-Based to Computer-Assisted
e. Experimental Mathematics: From Humans to Machines

11.1 Notation: From Rhetorical to Symbolic

We take symbols in mathematics for granted. Without a well-developed symbolic notation, mathematics would be inconceivable to us. We should note, however, that the subject evolved for at least three millennia with hardly any symbols! The historian of mathematics Kirsti Pederson notes the impact of the lack of notation on the early development of calculus [10, p. 47]:

> » An important reason why mathematicians [of the early seventeenth century] failed to see the general perspectives inherent in their various methods [for solving calculus problems] was probably the fact that to a great extent they expressed themselves in ordinary language without any special notation and so found it difficult to formulate the connections between the problems they dealt with.

In crucially important developments, symbolic notation was introduced in *algebra* by François Viète (1540–1603) in the late sixteenth century, and in *calculus* by Gottfried Leibniz (1646–1716) and Isaac Newton (1642–1727) in the late seventeenth century. Leibniz' superior notation prevailed over Newton's. Its pedagogical advantages are well expressed by the mathematician Charles Edwards:

> » It is hardly an exaggeration to say that the calculus of Leibniz [unlike that of Newton] brings within the range of an ordinary student problems that once required the ingenuity of an Archimedes or a Newton [9, p. 232].

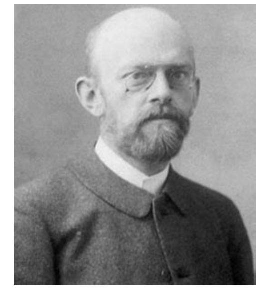
A good notation aids not only in the *proof* of results but also in their *discovery*. A poor notation can impede progress.

Two more examples of superb notations are Carl Friedrich Gauss' for congruences and Arthur Cayley's for matrices. Also important was the introduction of notations for positive

9. Franzén, T.: Gödel's Theorem: An Incomplete Guide to Its Use and Abuse. A K Peters, Natick (2005)
10. Gold, B., Simons, A. (eds): Proof & Other Dilemmas: Mathematics and Philosophy. Mathematical Association of America, Washington DC (2008)
11. Hankins, T.L.: Sir William Rowan Hamilton. Johns Hopkins University Press, Baltimore (1980)
12. Hersh, R.: What is Mathematics, Really? Oxford University Press, New York (1997)
13. Katz, V.: A History of Mathematics: An Introduction, 3rd edn. Addison-Wesley, Boston (2009)
14. Kline, M.: Mathematical Thought from Ancient to Modem Times. Oxford University Press, New York (1972)
15. Kline, M.: Mathematics: The Loss of Certainty. Oxford University Press, New York (1980)
16. Manin, Yu.I.: How convincing is a proof? Math. Intell. **2**(1), 17–18 (1979)
17. Moore, G.H.: Zermelo's Axion of Choice: Its Origins, Development, and Influence. Springer, New York (1982)
18. Rucker, R.: Infinity and the Mind: The Science and Philosophy of the Infinite. Birkhäuser, Boston (1982)
19. Stillwell, J.: Roads to Infinity: The Mathematics of Truth and Proof. A K Peters, Natick (2010)
20. von Neumann, J.: The role of mathematics in the sciences and in society. In: Taub, A.H. (ed.) Collected Works, Vol. 6, pp. 477–490. Macmillan, New York (1963)
21. Weyl, H.: Mathematics and logic. Am. Math. Mon. **53**, 2–13 (1946)

Further Reading

22. Bridges, D., Richman, F: Varieties of Constructive Mathematics. Cambridge University Press, Cambridge (1987)
23. Caldar, A.: Constructive mathematics. Sci. Am. **241**, 146–171 (1979)
24. Edwards, C.H.: The Historical Development of the Calculus. Springer, New York (1979)
25. Goodman, N.: Mathematics as an objective science. Am. Math. Mon. **86,** 540–551 (1979)
26. Kitcher, P.: The Nature of Mathematical Knowledge. Oxford University Press, New York (1983)
27. Mandelkern, M.: Constructive mathematics. Math. Mag. **58**, 272–280 (1985)
28. Richman, F.: Existence proofs. Am. Math. Mon. **106**, 303–308 (1999)
29. Rota, G.-C.: The phenomenology of mathematical proof. Synthese **111**, 183–196 (1997)
30. Tarski, A.: Truth and proof. Sci. Am. **220**, 63–77 (1969)
31. von Neumann, J.: The mathematician. In: Newman, J.R. (ed.) The World of Mathematics, Vol. 4, pp. 2053–2063. Simon and Schuster, New York (1956)
32. Weyl, H.: Axiomatic versus constructive procedures in mathematics. Math. Intell. **7**(4) 10–17 (1985)
33. Wilder, R.L.: The role of the axiomatic method. Am. Math. Mon. **74**, 115–127 (1967)
34. Wilder, R.L.: Evolution of Mathematical Concepts. Wiley, Hoboken (1968)



David Hilbert (1862–1943)

it was "the most influential book on geometry written in the [nineteenth] century" [1, p. 343], and E. T. Bell claimed that it "inaugurated the abstract mathematics of the 20th century" [1, p. 343].

1.3 The Modern Axiomatic Method

In the wake of Hilbert's *Foundations* one could define a "geometry" by picking a set of primitive terms—which, since it is to be a geometry, we might as well call "point", "line" …—and a consistent set of axioms, and logically deducing consequences from the axioms, which are then theorems of the geometry. Considerations like these gave rise in the nineteenth and early twentieth centuries to such geometries as desarguesian, nondesarguesian, finite, neutral, nonarchimedean, and inversive. Soon one began to describe (define) in Hilbert's manner mathematical structures other than geometries. Thus Giuseppe Peano defined (characterized) the positive integers in 1889 by means of the now classic Peano axioms, and Hilbert in 1900 gave a characterization of the real numbers as a complete ordered archimedean field. These accomplishments were in line with the spirit of rigor, generalization, abstraction, and axiomatization prevailing in late nineteenth- and early twentieth-century mathematics. Among early exponents of this approach were Dedekind, Peano, and especially Hilbert himself.

Yet another approach to axiomatics was begun in algebra and resulted in the now familiar algebraic structures of groups, rings, fields, vector spaces, modules, and ideals. These structures arose mainly from mathematicians' inability to solve old problems by old means, which necessitated the introduction of new structures. For the story of the rise of the group concept see [15]. Topological spaces, normed rings, Hilbert spaces, and lattices are among many other examples of mathematical structures defined by axiom systems. These structures, unlike (say) euclidean geometry, the natural numbers, or the real numbers, do not characterize a unique mathematical entity, but rather subsume many (usually infinitely many) different objects under the roof of a single set of axioms.

The rise of modern axiomatics—one of the most distinctive features of modern mathematics—was gradual and slow, lasting for much of the nineteenth century and the early decades of the twentieth. In the 1920s the axiomatic method became well established in a number of major areas of mathematics, including algebra, analysis, geometry, and topology, and it flourished

1

during the following three decades. Nicolas Bourbaki, among its most able practitioners and promoters, gave an eloquent description of the essence of the axiomatic method at what was perhaps the height of its power, in 1950 [2, p. 223]:

> What the axiomatic method sets as its essential aim, is exactly that which logical formalism by itself cannot supply, namely the profound intelligibility of mathematics. Just as the experimental method starts from the *a priori* belief in the permanence of natural laws, so the axiomatic method has its cornerstone in the conviction that, not only is mathematics not a randomly developing concatenation of syllogisms, but neither is it a collection of more or less "astute" tricks, arrived at by lucky combinations, in which purely technical cleverness wins the day. Where the superficial observer sees only two, or several, quite distinct theories, lending one another "unexpected support" through the intervention of a mathematician of genius, the axiomatic method teaches us to look for the deep-lying reasons for such a discovery, to find the common ideas of these theories, buried under the accumulation of details properly belonging to each of them, to bring these ideas forward and to put them in their proper light.

In this article Bourbaki presents a panoramic view of mathematics organized around what he calls "mother structures"—algebraic, ordered, and topological, and various substructures and cross-fertilizing structures. This must have been an alluring, even bewitching, perspective to those growing up mathematically during this period.

1.4 Ancient vs. Modern Axiomatics

There are significant differences between Euclid's axiomatics and its modern incarnation in the nineteenth and twentieth centuries. Comparing Euclid's *Elements* with Hilbert's *Foundations of Geometry* makes starkly clear how standards of rigor have evolved. Moreover, while the chief role played by the axiomatic method in ancient Greece was (probably) that of providing a sure foundation, it became in the first half of the twentieth century a tool of research. Note, for example, the rich and deep theory of groups, which comprises the logical consequences of a "simple" set of four axioms.

The modern axiomatic method was also indispensable in clarifying the status of various mathematical methods and results, such as the axiom of choice and the continuum hypothesis, to which mathematicians' intuition provided little guide. And it played an essential role in the *discovery* of certain concepts, results, and theories. For example, the desarguesian and non-desarguesian geometries "could never have been discovered without [the axiomatic] method" [4, p. 182]. Thus the sometimes opposed activities of discovery and demonstration coexisted within the axiomatic method.

The modern axiomatic method was however not universally endorsed. Although some, notably Hilbert, claimed that it is the central method of mathematical thought, others, for instance Klein, argued that as a method of discovery it tends to stifle creativity. And it has its limitations as a method of demonstration. The following quotation from Hermann Weyl (1885–1955) puts the issue in a broader perspective [13, p. 38; his italics]:

> *Large parts of modern mathematical research are based on a dexterous blending of axiomatic and constructive procedures.*

It is probably safe to say, however, that most mathematicians are untroubled, at least in their daily work, about the debates over their subject's underpinnings. "I think", says the distinguished modern mathematician Richard Askey, "there is far too much emphasis on […] the foundations of mathematics in much of what is published on the history of mathematics" [1, p. 203]. Philip Davis and Reuben Hersh put the issue in perspective [4, p. 318]:

> If you do mathematics every day, it seems the most natural thing in the world. If you stop to think about what you are doing and what it means, it seems one of the most mysterious.

Weyl says it more lyrically:

> The question of the ultimate foundations and the ultimate meaning of mathematics remains open; we do not know in what direction it will find its final solution or even whether a final objective answer can be expected at all. 'Mathematizing' may well be a creative activity of man, like language or music, of primary originality, whose historical decisions defy complete objective rationalization [15, p. 6].

Problems and Projects

1. What is Platonism, and how is it related to Plato's view of mathematics?
2. Discuss the axiom of choice. Why was it controversial?
3. Discuss briefly the Zermelo–Fraenkel axiomatization of set theory. How did it avoid Russell's paradox?
4. What is the continuum hypothesis? Discuss Gödel's and Cohen's results dealing with this hypothesis.
5. What are cantorian and noncantorian set theories? Compare with euclidean and noneuclidean geometries.
6. Discuss humanism, a philosophy of mathematics proposed by Reuben Hersh. See [12].
7. Discuss the philosophy of proof outlined by Imre Lakatos (1922–1974) in his booklet *Proofs and Refutations*.
8. It has been claimed that Gödel's incompleteness theorems imply the intellectual superiority of humans over machines. Discuss. See for example [9].
9. Discuss the role of proof in mathematics and changes in its practice. See [4, 5, 10, 14].

10

References

1. Aspray, W., Kitcher, P. (eds.): History and Philosophy of Modern Mathematics. University of Minnesota Press, Minneapolis (1988)
2. Bishop, E.: Foundations of Constructive Analysis. McGraw-Hill, New York (1967)
3. Davis, M.: Engines of Logic: Mathematicians and the Origins of the Computer. W.W. Norton & Co. New York (2000)
4. Davis, P.J., Hersh, R: The Mathematical Experience. Birkhäuser, Boston (1981) (Revised Study Edition, with E. Marchisotto, 1995)
5. De Millo, R.A., Lipton, R.J., Perlis, A.J.: Social processes and proofs of theorems and programs. Communications of the ACM **22**, 271–280 (1979)
6. Edwards, H.M.: Kronecker's algorithmic mathematics. Math. Intell. **31**(2), 11–14 (2009)
7. Eves, H.: Great Moments in Mathematics, 2 vols. (before 1650, and after 1650, resp.). Mathematical Association of America Washington DC (1983)
8. Feferman, S.: What does logic have to tell us about mathematical proofs? Math. Intell. **2**(1) 20–24 (1979)

Hermann Weyl (1885–1955)

nonconstructive proofs that they inform the world that a treasure exists without disclosing its location [14, p. 103].

On the other hand, the proofs of the intuitionists are certainly acceptable to the formalists. Many results in analysis, and more recently in algebra, have been reconstructed, thanks to the pioneering effort of Errett Bishop (1928–1983), using finitistic methods [2]. For example, as early as 1924 Brouwer and Weyl gave constructive proofs yielding a root of a complex polynomial—but actually *finding* such a root may require up to 10^{10} years! Yuri Manin, a prominent Russian mathematician, suggests that the mathematician "should at least be willing to admit that proof can have objectively different 'degrees of proofness'" [16, p. 17].

10.8 Conclusion

The differences between the formalists and the intuitionists on the one hand, and their nineteenth-century forerunners on the other hand, were genuine. For the first time, mathematicians were seriously and irreconcilably divided over what constitutes a proof in mathematics. Moreover, this division seems to have had an impact on the work that at least some mathematicians chose to pursue, as the testimony of two of the most prominent practitioners of that epoch—John von Neumann and Hermann Weyl, respectively—indicates:

> In my own experience… there were very serious substantive discussions as to what the fundamental principles of mathematics are; as to whether a large chapter of mathematics is really logically binding or not…. It was not at all clear exactly what one means by absolute rigor, and specifically, whether one should limit oneself to use only those parts of mathematics which nobody questioned. Thus, remarkably enough, in a large fraction of mathematics there actually existed differences of opinion! [20, p. 480].

> Outwardly it does not seem to hamper our daily work, and yet I for one confess that it has had a considerable practical influence on my mathematical life. It directed my interests to fields I considered relatively 'safe', and has been a constant drain on the enthusiasm and determination with which I pursued my research work [21, p. 13].

And finally, a comment from Bourbaki—a masterful practitioner and strong advocate of the axiomatic method [2, p. 231]:

> The unity which [the axiomatic method] gives to mathematics is not the armor of formal logic, the unity of a lifeless skeleton; it is the nutritive fluid of an organism at the height of its development, the supple and fertile research instrument to which all the great mathematical thinkers since Gauss have contributed, all those who, in the words of Lejeune-Dirichlet, have always labored to "substitute ideas for calculations".

Problems and Projects

1. Write a brief biography of Hilbert.
2. Describe Hilbert's characterization of the real numbers as a complete, ordered, archimedean field. What geometric purpose was it intended to serve?
3. Discuss several propositions in Euclid's *Elements* dealing with number theory.
4. Discuss several propositions in Book II of the *Elements* which correspond to algebraic results.
5. Skech a "proof", using axioms in Euclid's *Elements,* that every triangle is equilateral. Where is the error to be found? How is it to be corrected?
6. What was the fate of Euclid's five postulates in Hilbert's *Foundations of Geometry?* See for example [6].
7. Discuss Hilbert's "Axioms of Betweenness" and some of the uses he made of them.
8. Discuss Pasch's axioms for projective geometry.
9. Describe how the (algebraic) concept of a "ring" arose in the late nineteenth and the early twentieth centuries. See [7, 8].

References

1. Birkhoff, G., Bennett, M.K.: Hilbert's *Grundlagen der Geometrie*. Rend. Circ. Mat. Palermo. Ser. **II**(36), 343–389 (1987)
2. Bourbaki, N.: The architecture of mathematics. Am. Math. Mon. **57**, 221–232 (1950)
3. Eves, H.: Great Moments in Mathematics, 2 vols.; (a) before 1650, and (b) after 1650. Mathematical Association of America, (1983)
4. Gray, J.: Review of *Uber die Enstehung von David Hilberts 'Grundlagen der Geometrie'*. Hist. Math. **15**, 181–183 (1988)
5. Gray, J.: Worlds Out of Nothing: A Course in the History of Geometry in the 19th Century. Springer, Berlin (2007)
6. Greenberg, M.J.: Euclidean and Non-euclidean Geometries: Development and History, 22nd edn. W. H. Freeman, New York (1980)
7. Katz, V.: A History of Mathematics: An Introduction, 3rd. edn.. Addison-Wesley, Boston (2009)
8. Kline, M.: Mathematical Thought from Ancient to Modern Times. Oxford University Press, Oxford (1972)
9. Knorr, W.R.: On the early history of axiomatics: the interaction of mathematics and philosophy in Greek antiquity. In: Hintikka, J. et al. (eds.) Theory Change, Ancient Axiomatics and Galileo's Methodology, pp. 145–186. D. Reidel, Dordrecht (1980)
10. Lloyd, G.E.R.: Magic, Reason and Experience: Studies in the Origin and Development of Greek Science. Cambridge University Press, Cambridge (1979)
11. Szabo, A.: The Beginnings of Greek Mathematics. D. Reidel, Dordrecht (1978)
12. van der Waerden, B.L.: Science Awakening. Noordhoff, The Netherlands (1954)
13. Weyl, H.: Axiomatic versus constructive procedures in mathematics. Math. Intell. 7(4), 10–17, 38 (1985)

1

14. Wilder, R.L.: Evolution of Mathematical Concepts. Wiley, Hoboken (1968)
15. Wussing, H.: The Genesis of the Abstract Group Concept. MIT Press, Cambridge (1984). (Translated from the German by A. Shenitzer, with the editorial assistance of H. Grant)

Further Reading

16. Artmann, B.: Euclid: The Creation of Mathematics. Springer, Berlin (1999)
17. Cuomo, S.: Ancient Mathematics. Routledge, London (2001)
18. Davis, P.J., Hersh, R.: The Mathematical Experience. Birkhäuser, Boston (1981) (Revised Study Edition, with E. Marchisotto, published in 1995)
19. Mueller, I.: Philosophy of Mathematics and Deductive Structure in Euclid's Elements. MIT Press, Cambridge (1981)
20. Netz, R.: The Shaping of Deduction in Greek Mathematics: A Study in Cognitive History. Cambridge University Press, Cambridge (1999)
21. von Neumann, J.: The mathematician. In: Newman, J.R. (ed.) The World of Mathematics, Vol. 4, pp. 2053–2063. Simon and Schuster, New York (1956)
22. von Neumann, J.: The role of mathematics in the sciences and in society. In: Taub, A.H. (ed.) Collected Works, Vol. 6, pp. 477–490. Macmillan, London (1963)
23. Weyl, H.: Mathematics and logic. Am. Math. Mon. **53**, 2–13 (1946)
24. Wilder, R.L.: The role of the axiomatic method. Am. Math. Mon. **74**, 115–127 (1967)

Luitzen Egbertus Jan Brouwer (1881–1966)

10

Among the results unacceptable to the intuitionists is the law of trichotomy: Given any real number N, either $N > 0$ or $N = 0$ or $N < 0$. The following example substantiates that point [10, p. xx]:

Define a real number N as follows: $N = \sum_{n=2}^{\infty} \frac{a_n}{10^n}$, where

$$a_n = \begin{cases} 1, \text{if } 2n \text{ is the first even integer that is not the sum of two primes,} \\ n > 1,\ n \text{ even,} \\ -1, \text{ if } 2n \text{ is the first even integer that is not the sum of two primes.} \\ n > 1,\ n \text{ odd,} \\ 0, \text{ otherwise} \end{cases}$$

The definition of N is acceptable to both the formalists and the intuitionists; its digits can be calculated—at least in theory—to any degree of accuracy. But to the intuitionists, none of $N > 0$, $N < 0$, or $N = 0$ is meaningful since it is not known if Goldbach's conjecture (that every even number greater than 2 is a sum of two primes) is true or false. Thus the law of trichotomy fails.

10.7 Nonconstructive Proofs

A prominent feature of nineteenth-century mathematics was nonconstructive existence proofs—which were almost unknown before that time. Thus Gauss proved the "Fundamental Theorem of Algebra" about the existence of roots of polynomial equations without showing how to find them. Augustin-Louis Cauchy and others proved the existence of solutions of differential equations without providing the solutions explicitly. Cauchy proved the existence of the integral of an arbitrary continuous function, but was often unable to evaluate integrals of specific functions. He gave tests of convergence of series without indicating what they converge to. Late in the century Hilbert proved the existence of, but did not explicitly construct, a finite basis for any ideal in a polynomial ring. Richard Dedekind constructed the real numbers by using completed infinities. Such examples abound. All were rejected by the intuitionists. Hermann Weyl said of

Although Gödel's results are of fundamental *philosophical* consequence, they have not affected the daily work of most mathematicians, although "it is most likely safe to say that no mathematical theorem has aroused as much interest among nonmathematicians as Gödel's Incompleteness Theorem[s]" [9, p. 1].

10.5 Mathematics and Faith

Just as in the nineteenth century, following the invention of noneuclidean geometries, noncommutative algebras, and other developments, mathematics lost its claim to (absolute) truth (see ► Chapter 7), so in the twentieth century, following Gödel's incompleteness theorems, it lost its claim to certainty. In the nineteenth century truth in mathematics was replaced by validity (relative truth), and in the twentieth century certainty was replaced by faith. For a formal, twentieth-century notion of truth in mathematics and its relation to proof see [19].

Mathematics and faith? Surely the two are incompatible. But the mathematician Howard Eves (1911–2004) makes the case for such an association, on the following grounds. First, a definition of religion, by the mathematician Frank De Sua (1921–2013) [3, p. 305]:

> Religion is any discipline whose foundations rest on an element of faith, irrespective of any element of reason which may be present.

For example, quantum mechanics would be a religion under this definition. But, given Godel's result, that for many systems there are truths expressible in that system which are not provable, it follows that:

> Mathematics is the only branch of theology possessing a rigorous demonstration of the fact that it should be so classified [4, p. 305].

See [7].

10.6 Intuitionism

The intuitionists, headed by Luitzen Egbertus Jan Brouwer, claimed that no formal analysis of axiomatic systems is necessary. In fact, mathematics should not be founded on systems of axioms. The mathematicians' intuition, beginning with that of number, will guide them in avoiding contradictions. They must, however, pay special attention to definitions and methods of proof. These must be constructive and finitistic. In particular, the law of the excluded middle, completed infinities, the axiom of choice, and proof by contradiction are all outlawed. Hilbert protested that

> taking the principle of the excluded middle from the mathematician would be the same, say, as proscribing the telescope to the astronomer or to the boxer the use of his fists [15, p. 246].

Solution by Radicals of the Cubic: From Equations to Groups and from Real to Complex Numbers

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics,* Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_2,

2.1 Introduction

Several ancient civilizations—the Babylonian, Egyptian, Chinese, and Indian—dealt with the solution of polynomial equations, mainly linear. The Babylonians (ca. 1600 BC ff.) were particularly proficient "algebraists". They were able to solve quadratic equations, as well as equations that *lead to* quadratics—for example, $x + y = a$ and $x^2 + y^2 = b$—by methods similar to ours. The equations were given in the form of "word problems", and were often expressed in geometric language. Here is a typical example [7, p. 24]:

> I summed the area and two-thirds of my square-side and it was 0;35 [35/60 in sexagesimal notation]. [What is the side of my square?]

In modern notation, the problem is to solve the equation $x^2 + (2/3)x = 35/60$. See [7, p. 24] for the Babylonians' solution of this equation.

The Chinese (ca. 200 BC ff.) and the Indians (ca. 600 BC ff.)—in each case the dates are very rough—made considerable advances in algebra. For example, both allowed negative coefficients in their equations—though not negative roots—and admitted two roots for a quadratic equation. They also described procedures for manipulating equations. The Chinese had methods for approximating roots of polynomial equations of any degree, and they solved *systems* of linear equations using "matrices" (rectangular arrays of numbers) well before such techniques were developed in Western Europe. The mathematics of the ancient Greeks, in particular their geometry and number theory, was relatively advanced, but their algebra was rather weak. (Note however that Diophantus (fl. ca. 250 AD), in his great number-theoretic work *Arithmetica,* introduced various algebraic symbols [1].) Book II of Euclid's remarkable work *Elements* (ca. 300 BC) presents, in geometric language, results which are familiar to *us* as algebraic, but most modern scholars believe that the Greeks of this period were *not* thinking algebraically.

Islamic mathematicians made important contributions in algebra between the ninth and fifteenth centuries. Among the foremost was Muhammad ibn-Mūsā al-Khwārizmī, dubbed by some "the Euclid of algebra" because he systematized the subject as it then existed and made it into an independent field of study. He did this in his book *al-jabr w al-muqabalah*. "Al-jabr", from which stems our word "algebra", denotes the moving of a negative term of an equation to the other side so as to make it positive, and "al-muqabalah" refers to cancelling equal (positive) terms on the two sides of an equation. These are of course basic procedures for solving polynomial equations. al-Khwārizmī, from whose name is derived the word "algorithm", applied these procedures to the solution of quadratic equations, which he classified into five types: $ax^2 = bx$, $ax^2 = b$, $ax^2 + bx = c$, $ax^2 + c = bx$, and $ax^2 = bx + c$. This categorization was necessary since al-Khwārizmī did not admit negative coefficients or zero into the number system. He also had no algebraic notation, so that his problems and solutions were expressed rhetorically (in words). He did however offer (geometric) justification for his solution procedures.

2.2 Cubic and Quartic Equations

The Babylonians (as we mentioned) were solving quadratic equations by about 1600 BC, using essentially an equivalent of our "quadratic formula". A natural question was therefore whether *cubic* equations could be solved using "similar" formulas; three thousand years would pass before the answer was discovered. It was a great event in algebra when mathematicians of the sixteenth century succeeded in solving—by radicals—not only cubic but also quartic equations. This accomplishment was very much in character with the mood of the Renaissance—which wanted not only to absorb the classic works of the ancients but to strike out in new directions. Indeed, the solution of the cubic unquestionably proved a far-reaching departure.

A "solution by radicals" of a polynomial equation is a formula giving the roots of the equation in terms of its coefficients. The only permissible operations to be applied to the coefficients are the four algebraic operations (addition, subtraction, multiplication, and division) and the extraction of roots (square roots, cube roots, and so on, that is, "radicals"). For example, the quadratic formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ is a solution by radicals of the equation $ax^2 + bx + c = 0$.

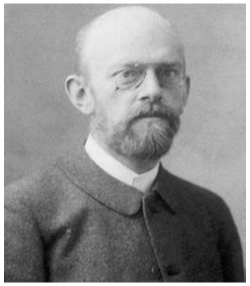
A solution by radicals of the cubic was first *published* in 1545 by Girolamo Cardano, in his *Ars Magna* (*The Great Art,* referring to algebra); it was *discovered* earlier by Scipione del Ferro and by Niccolò Tartaglia. The latter had passed on his method to Cardano, who had promised that he would not publish it; but he did. That is one version of events, which involved considerable drama and passion. A blow-by-blow account is given by Oysten Ore [12, pp. 53–107]. Here is Cardano's own rendition [7, p. 63]:

» Scipio Ferro of Bologna well-nigh thirty years ago [i.e., ca. 1515] discovered this rule and handed it on to Antonio Maria Fior of Venice, whose contest with Nicolò Tartaglia of Brescia gave Nicolò occasion to discover it. He [Tartaglia] gave it to me in response to my entreaties, though withholding the demonstration. Armed with this assistance, I sought out its demonstration in [various] forms. This was very difficult.

What came to be known as "Cardano's formula" for the solution of the cubic $x^3 = ax + b$ is given by

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\left(\frac{b}{2}\right)^2 - \left(\frac{a}{3}\right)^3}} + \sqrt[3]{\frac{b}{2} - \sqrt{\left(\frac{b}{2}\right)^2 - \left(\frac{a}{3}\right)^3}}.$$

David Hilbert (1862–1943)



Kurt Gödel (1906–1978)

10.4 Gödel's Incompleteness Theorems

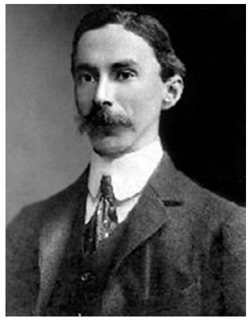
Hilbert's grand design was laid to rest by Kurt Gödel's two incompleteness theorems of 1931. These showed the inherent limitations of the axiomatic method:

1. The consistency of a large class of axiomatic systems, including those for arithmetic and set theory, cannot be established within the systems.
2. Moreover, any such system which is consistent must be incomplete. That is, given an axiomatic system, there will always be true results which are expressible in that system, but which cannot be established within the system.

For more technical statements, see [3, 8].

In connection with the first result, Weyl remarked: "God exists since mathematics is consistent and the devil exists since we cannot prove the consistency" [14, p. 1206]. (That mathematics is consistent is of course an article of *faith* of every working mathematician; see below.) The second result has elicited the (apparently anonymous) comment that Gödel gave a formal demonstration of the inadequacy of formal demonstrations.

Bertrand Russell (1872–1970)



fingers" [4, p. 272]. Moreover there were serious technical difficulties in the implementation of the logicist thesis. See [6, 13].

10.3 Formalism

The most serious debate within the mathematical community—it is still unresolved—has been between the adherents of the formalist and intuitionist schools. The formalist thesis, whose main exponent was Hilbert, views mathematics as a study of axiomatic systems. Both the primitive terms and the axioms of such a system are considered to be strings of symbols to which no meaning is to be attached. These are to be manipulated according to established rules of inference to obtain the theorems of the system.

At the time Hilbert advanced his thesis (the 1920s), the axiomatic method (see ▸ Chapter 1) had embraced much of algebra, arithmetic, analysis, set theory, and mathematical logic. Even though Ernst Zermelo's axiomatization of set theory in 1908 seemed to have avoided the paradoxes of that theory (see [17] and ▸ Chapter 9), there was no assurance that they would not resurface in one form or another. As Poincaré remarked [15, p. 1186]:

> We have put a fence around the herd to protect it from the wolves but we do not know whether some wolves were not already within the fence.

Hilbert felt that this possibility, and the denial of meaning to the primitive terms and postulates of axiomatic systems, made it imperative to undertake a careful analysis of such systems in order to establish their consistency. The methods by which this was to be accomplished were acceptable also to the intuitionists. These methods came to be known as "metamathematics" or "proof theory". See [3, 10].

The formalists have been accused of removing all meaning from mathematics and reducing it to symbol manipulation. Hilbert's aim, however, was to deal with the *foundations,* rather than with the daily practice of the mathematician. And to show that mathematics is free of inconsistencies one first needed to formalize the subject, the formalists claimed. This was formalism in the service of informality.

Girolamo Cardano (1501–1576)

See for example [1, 2, 5]. Several comments are in order:

i. Cardano used essentially no symbols, so his "formula" giving the solution of the cubic was expressed rhetorically.
ii. He was usually content with determining a single root of a cubic. But in fact, if a proper choice is made of the cube roots involved, all three roots of the equation can be determined from his formula.
iii. The coefficients and roots of the cubics he considered were specific *positive numbers,* so that he viewed (say) $x^3 = ax + b$ and $x^3 + ax = b$ as distinct. He devoted a chapter to the solution of each, and gave *geometric* justifications [13, p. 63 ff.].
iv. Negative numbers are found occasionally in his work, but he mistrusted them, and called them "fictitious". Irrational numbers were admitted as roots.

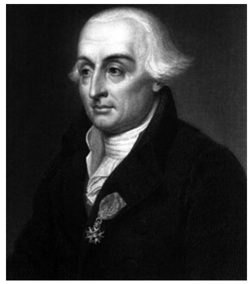
The solution by radicals of polynomial equations of the fourth degree—*quartics*—soon followed. The key idea was to reduce the solution of a quartic to that of a cubic. Ludovico Ferrari was the first to solve such equations, and his work was included in Cardano's *The Great Art* [4].

It should be pointed out that cubic equations had arisen—in geometric guise—already in ancient Greece (ca. 400 BC), in connection with the problem of trisecting an angle, and that methods for finding *approximate* roots of cubics and quartics were known, for example by Chinese and Moslem mathematicians, well before such equations were solved by radicals. The latter solutions, though exact, were of little *practical* value. But the ramifications of these "impractical" ideas were very significant, and will now be briefly sketched.

2.3 Beyond the Quartic: Lagrange

Having solved the cubic and quartic by radicals, mathematicians turned to finding a solution by radicals of the quintic (degree-five polynomial)—a quest that would take nearly 300 years. Some of the most distinguished mathematicians of the seventeenth and eighteenth centuries, among them François Viète, René Descartes, Gottfried Wilhelm Leibniz, Leonhard Euler, and Étienne Bezout, tackled the problem. The strategy was to seek new approaches to the solutions of the cubic and quartic, in the hope that at least one of them would generalize to the quintic.

Joseph Louis Lagrange (1736–1830)



But to no avail: although new ideas for solving the cubic and quartic *were* found, they did not yield the desired extensions. One approach, however, undertaken by Joseph Louis Lagrange in a paper of 1770 entitled *Reflections on the Algebraic Solution of Equations*, proved promising. Lagrange analyzed the various methods devised by his predecessors for solving cubic and quartic equations, and saw that—since those methods did not work when applied to the quintic—a deeper scrutiny was required. In his own words [14, p. 127]:

» I propose in this memoir to examine the various methods found so far for the algebraic solution of equations, to reduce them to general principles, and to let us see *a priori* why these methods succeed for the third and fourth degree, and fail for higher degrees.

Here are some of the key ideas of Lagrange's approach. With each polynomial equation of arbitrary degree n he associated a "resolvent equation", as follows: let f(x) be the original equation, with roots $x_1, x_2, x_3, \ldots, x_n$. (As is the usual practice, we denote by "f(x)" both the polynomial and the polynomial equation.) Pick a rational function $R(x_1, x_2, x_3, \ldots, x_n)$ of the roots and coefficients of f(x). (Lagrange described a method for doing this.) Consider the different values which $R(x_1, x_2, x_3, \ldots, x_n)$ assumes under all the n! permutations of the roots $x_1, x_2, x_3, \ldots, x_n$ of f(x). If these values are denoted by $y_1, y_2, y_3, \ldots, y_k$, the "resolvent equation" is $(x-y_1)(x-y_2)\ldots(x-y_k)$. Lagrange showed that k divides n!—the source of what we call "Lagrange's theorem" in group theory.

For example, if f(x) is a quartic with roots x_1, x_2, x_3, x_4, then $R(x_1, x_2, x_3, x_4)$ may be taken to be $x_1x_2 + x_3x_4$, and this function assumes three distinct values under the 24 permutations of x_1, x_2, x_3, and x_4. Thus, the resolvent equation of a quartic is a cubic. However, in carrying over this analysis to the quintic, Lagrange found that the resolvent equation is of *degree six,* rather than the hoped-for degree four.

Although Lagrange did not succeed in settling the problem of the solvability of the quintic by radicals, his work was a milestone. It was the first time that an association was made between the solutions of a polynomial equation and the permutations of its roots. In fact, Lagrange speculated that the study of the permutations of the roots of an equation was the cornerstone of the theory of algebraic equations—"the genuine principles of the solution of equations", as he put it [14, p. 146]. He was of course vindicated in this by Evariste Galois.

I believe that the numbers and functions of analysis are not the arbitrary product of our minds; I believe that they exist outside of us with the same character of necessity as the objects of objective reality; and we find or discover them and study them as do the physicists, chemists and zoologists [14, p. 1035].

The above three quotations, from Charles Hermite in 1893, Henri Poincaré in 1899, and again Hermite in 1905, respectively, are a reaction to various examples of "pathological" functions introduced during the previous half-century: integrable functions with discontinuities dense in any interval, continuous nowhere-differentiable functions, nonintegrable functions that are limits of integrable functions, and others.

» Later generations will regard *Mengenlehre* [Set Theory] as a disease from which one has recovered [14, p. 1003].

This is Poincaré again, speaking (in 1908) about Cantor's creation of set theory, especially in connection with the paradoxes that had arisen in the theory [17, 18]. Compare Poincaré's position with that of David Hilbert, the other giant of this period:

» No one shall expel us from the paradise which Cantor created for us [14, p. 1003].

The above sentiments, expressed by some of the leading mathematicians of the period, suggest an impending crisis. Although mathematical controversies had arisen before the nineteenth century, for example the vibrating-string controversy between d'Alembert and Euler, these were isolated cases. The frequency and intensity of the disaffection expressed in the nineteenth century were unprecedented and could no longer be ignored. The result was a split among mathematicians concerning the way they viewed their subject—its nature, meaning, and methods. The formal expression of that split was the rise in the early twentieth century of three schools of mathematical thought, three philosophies of mathematics—logicism, formalism, and intuitionism.

10.2 Logicism

The logicist thesis, expounded in the monumental *Principia Mathematica* of Bertrand Russell and Alfred North Whitehead, held that mathematics is part of logic. Mathematical concepts are expressible in terms of logical concepts; and mathematical theorems are tautologies, true by virtue of their form rather than their factual content. This thesis was motivated, in part, by the paradoxes in set theory, by the work of Gottlob Frege on mathematical logic and the foundations of arithmetic, and by the espousal of mathematical logic by Giuseppe Peano and his school. Its broad aim was to provide a foundation for mathematics. Although the logicist thesis was important philosophically and inspired subsequent work in mathematical logic, it was not embraced by the mathematical community. For one thing, it did not grant reality to mathematics other than in terms of logical concepts. For another, it took "forever" to obtain results of any consequence; for example, it is only on p. 362 of the *Principia* that Russell and Whitehead show that 1 + 1 = 2 (!); see [4, p. 334]. "If the mathematical process were really one of strict, logical progression", observe Richard A. De Millo (1947–) et al., "we would still be counting on our

Philosophy of Mathematics: From Hilbert to Gödel

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics,* Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_10, © Springer Science+Business Media, LLC 2015

10.1 Introduction

The nineteenth century witnessed a gradual transformation of mathematics—in fact, a gradual revolution, if that is not a contradiction in terms. Mathematicians turned more and more for the genesis of their ideas from the sensory and empirical to the intellectual and abstract. Although this subtle change already began in the sixteenth and seventeenth centuries with the introduction of such nonintuitive concepts as negative and complex numbers, instantaneous rates of change, and infinitely small quantities, these were often used (successfully) to solve physical problems and thus elicited little demand for justification.

But in the nineteenth century the introduction of noneuclidean geometries, noncommutative algebras, continuous nowhere-differentiable functions, space-filling curves, n-dimensional geometries, completed infinities of different sizes, and the like, could no longer be justified by physical utility. Georg Cantor's dictum that "the essence of mathematics is its freedom" [4, p. 448] became a reality—but one to which many mathematicians took strong exception, as the following quotations indicate (see ► Chapters 8, 9, 11).

> » There is still something in the system [of quaternions] which gravels me. I have not yet any clear view as to the extent to which we are at liberty arbitrarily to create imaginaries and to endow them with supernatural properties [11, p. 300].

The reservations are those of John Graves, who communicated them to his friend William Rowan Hamilton in 1844, shortly after the latter had invented the quaternions. The "supernatural properties" referred to the noncommutativity of multiplication of the quaternions.

> » Of what use is your beautiful investigation regarding π? Why study such problems since irrational numbers are nonexistent? [14, p. 1198] (But see [6, p. 13].)

This was Leopold Kronecker's damning praise of Carl Louis Ferdinand Lindemann, who proved in 1882 that π is transcendental, hence that the circle cannot be squared using straightedge and compass.

> » I turn away with fright and horror from this lamentable evil of functions without derivatives [14, p. 973].
>
> Logic sometimes makes monsters. For half a century we have seen a mass of bizarre functions which appear to be forced to resemble as little as possible honest functions which serve some purpose [14, p. 973].

Evariste Galois (1811–1832)

2.4 Ruffini, Abel, Galois

Paolo Ruffini and Niels-Henrik Abel proved (in 1799 and 1826, respectively) the unsolvability by radicals of the "general quintic". In fact, they proved the unsolvability of the "general equation" of degree n for *every* $n > 4$. They did this by building on Lagrange's pioneering ideas on resolvents. Lagrange had shown that a necessary condition for the solvability of the *general* polynomial equation of degree n is the existence of a resolvent of degree less than n. (A "general equation" is an equation with *arbitrary literal coefficients.*) Ruffini and Abel showed that such resolvents *do not exist* for *any* $n > 4$. (Abel proved this result without knowing of Ruffini's work; in any case, Ruffini's work had a significant gap.)

Although the *general* polynomial equation of degree > 4 is unsolvable by radicals, some specific equations of this form *are* solvable; for example, $x^n - 1 = 0$ is solvable by radicals for every $n > 4$. Galois characterized those equations that are solvable by radicals in terms of *group theory*: A polynomial is solvable by radicals if and only if its "Galois group" is "solvable". To prove this result Galois founded the elements of permutation group theory and introduced in it various important concepts, such as Galois group, normal subgroup, and solvable group. Thus ended, in the early 1830s, the great saga—beginning with Cardano in 1545—of solvability by radicals of equations of degrees greater than 2.

2.5 Complex Numbers: Birth

A hugely important development arising from the solution of the cubic by radicals was the introduction of *complex numbers*.

Recall that Cardano's solution of the cubic $x^3 = ax + b$ is given by

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\left(\frac{b}{2}\right)^2 - \left(\frac{a}{3}\right)^3}} + \sqrt[3]{\frac{b}{2} - \sqrt{\left(\frac{b}{2}\right)^2 - \left(\frac{a}{3}\right)^3}}.$$

Consider the cubic $x^3 = 9x + 2$. Its solution, using the above formula, is

$$x = \sqrt[3]{\tfrac{2}{2}+\sqrt{(\tfrac{2}{2})^2-(\tfrac{9}{3})^3}} + \sqrt[3]{\tfrac{2}{2}-\sqrt{(\tfrac{2}{2})^2-(\tfrac{9}{3})^3}} = \sqrt[3]{1+\sqrt{-26}} + \sqrt[3]{1-\sqrt{-26}}.$$

What is one to make of this solution? Since Cardano was suspicious of negative numbers—calling them “fictitious” [10, p. 40]—he certainly had no taste for their square roots, which he named “sophistic negatives” [10, p. 40]. He therefore regarded his formula as inapplicable to equations such as $x^3 = 9x + 2$. Judged by past experience, this was not an unreasonable attitude. For example, to pre-Renaissance mathematicians the quadratic formula could not be applied to $x^2 + 1 = 0$.

All this changed when the Italian Rafael Bombelli came on the scene. In his important book *Algebra* (1572) he applied Cardano's formula to the equation $x^3 = 15x + 4$ and obtained $x = \sqrt[3]{2+\sqrt{-121}} + \sqrt[3]{2-\sqrt{-121}}$. But he could not dismiss this solution, unpalatable as it would have been to Cardano, for he noted—by inspection—that $x = 4$ is also a root of this equation; its other two roots, $-2 \pm \sqrt{3}$, are also real numbers. Here was a paradox: while all three roots of the cubic $x^3 = 15x + 4$ are real, the formula used to obtain them involved square roots of negative numbers—meaningless at the time. How was one to resolve the paradox?

Bombelli had a “wild thought”: since the radicands $2+\sqrt{-121}$ and $2-\sqrt{-121}$ differ only in sign, the same might be true of their cube roots. He thus let

$$\sqrt[3]{2+\sqrt{-121}} = a + b\sqrt{-1}, \quad \sqrt[3]{2-\sqrt{-121}} = a - b\sqrt{-1},$$

and proceeded to solve for a and b by *manipulating these expressions according to the established rules for real variables*. He deduced that $a = 2$ and $b = 1$ and thereby showed that, indeed,

$$x = \sqrt[3]{2+\sqrt{-121}} + \sqrt[3]{2-\sqrt{-121}} = (2+\sqrt{-1}) + (2-\sqrt{-1}) = 4.$$

Bombelli had given meaning to the “meaningless”. He put it thus [11, p. 19]:

> It was a wild thought in the judgment of many; and I too for a long time was of the same opinion. The whole matter seemed to rest on sophistry rather than on truth. Yet I sought so long, until I actually proved this to be the case.

Moreover, Bombelli developed a “calculus” for complex numbers, stating such rules as $(+\sqrt{-1})(+\sqrt{-1}) = -1$ and $(+\sqrt{-1})(-\sqrt{-1}) = 1$, and defined addition and multiplication of some of these numbers. These innovations signaled the birth of complex numbers.

But note that this is a *retrospective* view of what Bombelli had done. He did not postulate the existence of a system of numbers—called complex numbers—containing the real numbers and satisfying basic properties of numbers. To him, the expressions he worked with were just that; they were important because they “explained” hitherto inexplicable phenomena. *Square roots of negative numbers could be manipulated in a meaningful way to yield significant results*. This was a bold idea indeed. See [8, 10].

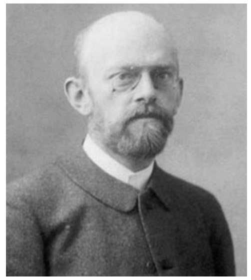
References

1. Bochner, S.: Infinity. In: Dictionary of the History of Ideas, II. Scribner's, New York (1973)
2. Bottazzini, U.: The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass. Springer-Verlag, New York (1986)
3. Boyer, C.B., Merzbach, U.C.: A History of Mathematics, 2nd edn. Wiley, New York (1989).
4. Dantzig, T.: Number: The Language of Science, 4th edn. Pi Press, New Jersey (2005) (orig. 1930).
5. Dauben, J.W.: The trigonometric background to Georg Cantor's theory of sets. Arch. Hist. Exact Sci. **7**, 181–216 (1970/1971)
6. Dauben, J.W.: Georg Cantor: His Mathematics and Philosophy of the Infinite. Harvard University Press, Cambridge (1979)
7. Dauben, J.W.: Georg Cantor and the origins of transfinite set theory. Sci. Am. **248**, 122–131 (1983)
8. Eves, H.: Great Moments in Mathematics (after 1650). Mathematical Association of America, Washington DC (1981)
9. Halmos, P.R.: Naïve Set Theory. Springer, New York (1974) (orig. 1960)
10. Kline, M.: Mathematical Thought from Ancient to Modern Times. Oxford University Press, New York (1972)
11. Maor, E.: To Infinity and Beyond: A Cultural History of the Infinite. Birkhäuser, Boston (1987)
12. Rucker, R.: Infinity and the Mind: The Science and Philosophy of the Infinite. Birkhäuser, Boston (1982)
13. Stillwell, J.: The continuum problem. Am. Math. Mon. **109**, 286–297 (2002)
14. Stillwell, J.: Roads to Infinity. A K Peters (2010)
15. Van Vleck, E.B.: The influence of Fourier series upon the development of mathematics. Science **39**, 113–124 (1914)

Further Reading

16. Davenport, A.A.: The Catholics, the Cathars, and the concept of infinity in the thirteenth century. Isis **88**, 263–295 (1997)
17. Vilenkin, N.Ya.: In Search of Infinity. Birkhäuser, Boston (1995) (Translated from the Russian by A. Shenitzer.)

David Hilbert (1862–1943)



There is much to make sense of in the above statements. We will address in ▸ Chapter 10 some of the issues that arise. For now, let us point out that, with the appropriate correctives, set theory—the study of the transfinite, the completed infinite—is alive and well. Its creation, initially at the hands of Cantor, followed by many other brilliant mathematicians, is one of the great turning points in the history of mathematics. It is important for students to internalize the following implications (among others) of transfinite arithmetic—all so contrary to their past experience.

9

1. The infinite is not merely the absence of the finite, nor is it merely "a manner of speaking". It is a precise mathematical concept, giving rise to a rich and deep theory.
2. The whole need not be greater than its parts. In fact, the whole is equal to infinitely many of its parts.
3. There are arithmetics which disobey such fundamental laws as cancellation under addition and multiplication, as well as the commutative laws of addition and multiplication; see "Problems and Projects", no. 8.
4. Infinity comes in different sizes—in fact, in infinitely many different sizes.

Problems and Projects

1. Describe the Achilles, Arrow, and Stadium paradoxes, proposed by Zeno.
2. Describe briefly the philosophy of Parmenides.
3. Discuss the impact of Zeno's paradoxes on Greek mathematics.
4. Describe some of the theological and philosophical discussions related to the infinite. See [4, 6, 12, 13].
5. Discuss briefly Cantor's life and some of his work (other than that discussed in this chapter).
6. Describe how Cantor's work on trigonometric series gave rise to his interest in the infinite. See [2, 5–7, 15].
7. Show that every infinite set contains a proper subset having the same cardinality as the original set. See [9].
8. What we have outlined in this chapter is the theory of cardinal numbers. There is a parallel, and arguably equally important, theory of *ordinal* numbers. In ordinal arithmetic, for example, $1+\omega \neq \omega+1$ for the ordinal ω (what is it?). Investigate the rudiments of ordinal arithmetic. See [3, 6, 7, 9, 12, 14].

Rafael Bombelli (1526–1572)

The equation $x^3 = 15x + 4$ considered above is an example of an "irreducible cubic", one with rational coefficients, irreducible over the rationals, all of whose three roots are real and distinct. It was shown in the nineteenth century that *any* solution by radicals of such a cubic—not just Cardano's—*must* involve complex numbers [5, 10]. Thus complex numbers are *unavoidable* when determining solutions by radicals of irreducible cubics. It is for this reason that they arose in connection with the solution of *cubic* rather than (as seems much more plausible) *quadratic* equations. Note that the nonexistence of a solution of the quadratic $x^2 + 1 = 0$ was accepted for centuries.

2.6 Growth

Here are several examples of the penetration of complex numbers into mathematics in the centuries after Bombelli.

As early as 1620, Albert Girard suggested that an equation of degree n *may* have n roots. Such statements of the "Fundamental Theorem of Algebra" were however vague and unclear. For example Descartes, who coined the unfortunate term "imaginary" for the new numbers (Gauss called them "complex"), stated that although one can imagine that every equation has as many roots as is indicated by its degree, no (real) numbers correspond to some of these imagined roots.

Leibniz, who spent considerable time and effort on the question of the meaning of complex numbers and the possibility of deriving reliable results by applying the ordinary laws of algebra to them, thought of complex roots as "an elegant and wonderful resource of divine intellect, an unnatural birth in the realm of thought, almost an amphibium between being and non-being" [11, p. 159].

Complex numbers were used by Johann Heinrich Lambert for map projection, by Jean le Rond d'Alembert in hydrodynamics, and by Euler, d'Alembert, and Lagrange in (incorrect) proofs of the Fundamental Theorem of Algebra.

Euler made important use of complex numbers in, for example, number theory and analysis; he also linked the exponential and trigonometric functions and, arguably, the five most important numbers in mathematics in, respectively, the following two famous formulas: $e^{ix} = \cos x + i \sin x$ and $e^{\pi i} + 1 = 0$. (Euler was the first to designate $\sqrt{-1}$ by "i".) Yet he said of them [9, p. 594]:

2

» Because all conceivable numbers are either greater than zero, less than zero or equal to zero, then it is clear that the square roots of negative numbers cannot be included among the possible numbers. Consequently we must say that these are impossible numbers. And this circumstance leads us to the concept of such numbers, which by their nature are impossible, and ordinarily are called imaginary or fancied numbers, because they exist only in the imagination.

Even the great Gauss, who in his doctoral thesis of 1797 gave the first essentially correct proof of the Fundamental Theorem of Algebra, claimed as late as 1825 that "the true metaphysics of $\sqrt{-1}$ is elusive" [9, p. 631]. But by 1831 Gauss had overcome these metaphysical scruples and, in connection with a work on number theory, published his scheme for representing them geometrically, as points in the plane. Similar representations by Caspar Wessel in 1797 and by Jean Robert Argand in 1806 had gone largely unnoticed; but when given Gauss' stamp of approval the geometric representation dispelled much of the mystery surrounding complex numbers.

Doubts concerning the meaning and legitimacy of complex numbers persisted for two and a half centuries following Bombelli's work. Yet during that same period these numbers were used extensively. *How can inexplicable things be so useful?* This is a recurrent theme in the history of mathematics. Bombelli's resolution of the paradox dealing with the solution of the cubic $x^3 = 15x + 4$ is an excellent example of this phenomenon.

2.7 Maturity

In the next two decades further developments took place. In 1833 William Rowan Hamilton gave an essentially rigorous algebraic definition of complex numbers as pairs of real numbers, and in 1847 Augustin-Louis Cauchy gave a completely rigorous definition in terms of congruence classes of real polynomials modulo $x^2 + 1$. In this he modelled himself on Gauss' definition of "congruences" for the integers. By the latter part of the nineteenth century most vestiges of mystery and distrust around complex numbers could be said to have disappeared [6].

But this is far from the end of their story. Various developments in mathematics in the nineteenth century gave us deeper insight into the role of complex numbers in mathematics and in other areas. These numbers offer just the right setting for dealing with many problems in mathematics in such diverse areas as algebra, analysis, geometry, and number theory. They have a symmetry and completeness that is often lacking in the real numbers. The following three quotations, by Gauss in 1811, Riemann in 1851, and Jacques Hadamard in the 1890s, respectively, say it well:

» Analysis … would lose immensely in beauty and balance and would be forced to add very hampering restrictions to truths which would hold generally otherwise, if … imaginary quantities were to be neglected [3, p. 31].

The original purpose and immediate objective in introducing complex numbers into mathematics is to express laws of dependence between variables by simpler operations on the quantities involved. If one applies these laws of dependence in an extended context, by giving the variables to which they relate complex values, there emerges a regularity and harmony which would otherwise have remained concealed [6, p. 64].

The shortest path between two truths in the real domain passes through the complex domain [9, p. 626].

Henri Poincaré (1854–1912)

$n < 2^n$, we conjecture that $c < 2^c$ = the number of subsets of R, and, more generally, that for any set A, $|A| < |P(A)|$, where P(A) is the set of all subsets of A, called the "power set" of A. It is easy to see that $|P(A)|$ is the number of functions $f: A \to \{0, 1\}$.

To show that $|A| < |P(A)|$, note first that $|A| \le |P(A)|$. If $|A| = |P(A)|$, then there exists a map $f: A \to P(A)$ which is 1–1 and onto. Let $B = \{b \in A : b \notin f(b)\}$. Since f is onto, pick $a \in A$ such that $f(a) = B$. Then $a \in B$ if and only if $a \notin B$, a contradiction. Thus $|A| < |P(A)|$.

Note that the "operator" P can be iterated, so that we get an infinite chain of increasing cardinal numbers, $|A| < |P(A)| < |P(P(A))| < \ldots$. This seemingly blissful state of affairs leads to serious difficulties. For if S = {all sets}, then for every set T, $|T| \le |S|$. In particular, $|P(S)| \le |S|$. But we have shown that $|A| < |P(A)|$ for any set A. So $|S| < |P(S)|$. This is a contradiction—a paradox. It is a serious paradox, for it mandates a restricted notion of set. In particular, S = {all sets} is not a set, although one would have thought that *any* collection of objects is a set. S is simply too large. So we need to restrict the notion of set (see ► Chapter 10).

The other question left open is whether there is a cardinal greater than $\aleph_0$ and less than c. Cantor thought there is none, and tried to prove it, without success. It turns out that both "yes" and "no" are valid answers! This mysterious statement must be clarified, of course; it *will* be, in ► Chapter 10.

9.9 Conclusion

Other problems and paradoxes arose in the theory of sets in the decades following Cantor's work. For example, consider the set $S = \{x : x \notin x\}$. Then $S \in S$ if and only if $S \notin S$. This is the famous Russell Paradox. A mathematical school arose which viewed the completed infinite as taboo, as Aristotle had urged more than two millennia earlier. The potential infinite will do just fine, that school argued. We can recover much of known mathematics without its use, they claimed. Poincaré, one of its outstanding early founders, declared that "later generations will regard [Cantor's] set theory as a disease from which one has recovered" [10, p. 1003]. But this was a minority opinion. Hilbert, representing the majority view, countered: "No one shall expel us from the paradise which Cantor has created for us" [10, p. 1003].

9.7 Arithmetic

We can now perform—informally—some transfinite arithmetic. Note that we define the sum and product of cardinals as follows: If A and B are two sets, define $|A|+|B|=|A\cup B|$ and $|A|\times|B|=|A\times B|$, picking, without loss of generality, A and B so that $A\cap B=\varnothing$ ($A\times B$ is the Cartesian product of A and B, $\varnothing$ is the empty set).

Since both the natural numbers and the nonnegative integers are denumerable, $\aleph_0=1+\aleph_0$; also $\aleph_0=2+\aleph_0$, so that $1+\aleph_0=2+\aleph_0$. Hence, by cancellation, $1=2$. Another paradox? Not quite. Of course we should not expect the laws of transfinite arithmetic to be identical to those of our daily arithmetic of finite numbers. In particular, the cancellation law under addition is invalid in transfinite arithmetic.

We can show by induction that $n+\aleph_0=\aleph_0$. It is also easy to see that $\aleph_0+\aleph_0=\aleph_0$, since, for example, $N=E\cup O$, where each of the natural numbers (N), the even positive integers (E), and the odd positive integers (O) has cardinality $\aleph_0$. As for multiplication, note that $2\times\aleph_0=\aleph_0+\aleph_0=\aleph_0$, so that $1\times\aleph_0=2\times\aleph_0$. Thus cancellation under multiplication is also invalid in transfinite arithmetic. We also have $n\times\aleph_0=\aleph_0$, by induction, and $\aleph_0\times\aleph_0=\aleph_0$, since the rationals are pairs of integers. So $\aleph_0{}^2=\aleph_0$, and by induction, $\aleph_0{}^n=\aleph_0$.

If we denote the cardinality of the real numbers by c (the continuum), we have shown that $\aleph_0<c$. Since every line segment has cardinality c, it follows that $c+c=c$, and since the real and the complex numbers have the same cardinality, $c\times c=c$. Now, $c\le\aleph_0+c<c+c=c$, so that $\aleph_0+c=c$. Similarly, $\aleph_0\times c=c$.

Two important problems that Cantor had to contend with were to determine the cardinalities of the irrational and the transcendental numbers. We show that they are both c. Denote the transcendental numbers by T and let $t=|T|$; then $c=t+\aleph_0=|T|+|A|$, where A are the algebraic numbers. Now "peel off" $\aleph_0$ elements from T, say $T=J\cup K$, where $|K|=\aleph_0$. Then $c=|T|+|A|=|J\cup K|+|A|=|J|+|K|+|A|=|J|+|K|=|T|=t$, so that $t=c$. Now, if I denotes the irrationals, then, since $T\le I$, $c=|T|\le|I|\le|\mathbf{R}|=c$, so $|I|=c$ (**R** denotes the real numbers).

Transcendental numbers were defined by Euler in the eighteenth century, but the first example of such numbers was given by Joseph Liouville in 1844. Charles Hermite proved in 1873 that e is transcendental, and Carl Louis Ferdinand Lindemann showed in 1882 that the same is true of π. (This last result implies that the circle cannot be squared.) Cantor (in 1874) proved the remarkable fact that there are *more* transcendental numbers than algebraic numbers. But he did not actually *construct* any transcendental numbers—his was an example of a "nonconstructive" existence proof. Such proofs were rejected by a number of very eminent mathematicians, among them Kronecker, Poincaré, Weyl, and Brouwer (see ▶ Chapter 10).

9.8 Two Major Problems

The inequality $\aleph_0<c$ posed two major problems for Cantor: are there cardinals between $\aleph_0$ and c, and are there any beyond c? The second question is relatively easy to answer.

We take our cue from the finite case. Note that $2^n=(1+1)^n=\sum_{i=0}^{i=n}\binom{n}{i}=$ the number of subsets of $\{1,2,3,\ldots,n\}=$ the number of functions $f:\{1,2,3,\ldots,n\}\to\{0,1\}$. Since

We give brief indications of what is involved in welcoming complex numbers into mathematics.

In algebra, their introduction gave us the celebrated "Fundamental Theorem of Algebra": every equation with complex coefficients has a complex root. The complex numbers offer an example of an "algebraically closed field", relative to which many problems in linear algebra and other areas of abstract algebra have their "natural" formulation and solution.

In analysis, the nineteenth century saw the development of a powerful and beautiful branch of mathematics: "complex function theory". One indication of its efficacy is that a function in the complex domain is infinitely differentiable if once differentiable—which of course is false for functions of a real variable.

In geometry, the complex numbers lend symmetry and generality to the formulation and description of its various branches, including euclidean, inversive, and noneuclidean geometry. For a specific example we mention Gauss' use of complex numbers to show that the regular polygon of seventeen sides is constructible with straightedge and compass.

In number theory, certain diophantine equations can be solved using complex numbers. For example, the domain consisting of the set of elements of the form $a+b\sqrt{2}\,i$, with *a* and *b* integers, has *unique factorization,* and in it the Bachet equation $x^2+2=y^3$ factors as $(x+\sqrt{2}\,i)(x-\sqrt{2}\,i)=y^3$. This greatly facilitates its solution (in integers).

An elementary illustration of Hadamard's dictum that "the shortest path between two truths in the real domain passes through the complex domain" is supplied by the following proof that the product of sums of two squares of integers is again a sum of two squares of integers; that is, given integers a, b, c, and d, there exist integers u and v such that $(a^2+b^2)(c^2+d^2)=u^2+v^2$. For, $(a^2+b^2)(c^2+d^2)=(a+bi)(a-bi)(c+di)(c-di)$ $=[(a+bi)(c+di)][(a-bi)(c-di)]=(u+vi)(u-vi)=u^2+v^2$ for some integers u and v. Try to prove this result without the use of complex numbers and without being given the u and v in terms of a, b, c, and d.

In addition to their fundamental uses in mathematics, complex numbers have become indispensible in science and technology. For example, they are used in quantum mechanics and in electric circuitry. The "impossible" has become not only possible but essential [6].

Problems and Projects

1. Discuss the solution of the quartic by radicals.
2. Research the lives and work of two mathematicians discussed in this chapter.
3. Show how to trisect an angle using trigonometric functions.
4. Discuss the Italian Renaissance, including some of its accomplishments in mathematics (those not discussed in this chapter).
5. Describe the "geometric algebra" of the ancient Greeks.
6. Discuss the algebra of al-Khwārizmī.
7. Show how to solve an elementary problem in euclidean geometry using complex numbers.
8. Discuss the meaning of the logarithms of negative and complex numbers.
9. The "quaternions" (also known as "hypercomplex numbers") contain the complex numbers. Discuss some of their properties that are like those of the complex numbers and some that differ.
10. Show how to resolve the paradox of the irreducible cubic $x^3=15x+4$.

References

1. Bashmakova, I.G., Smirnova, G.S.: The beginnings and evolution of algebra. Mathematical Association of America (2000). (Translated from the Russian by A. Shenitzer)
2. Berlinghoff, W.P., Gouvêa, F.Q.: Math through the ages: A gentle history for teachers and others, expanded edition. Mathematical Association of America (2004)
3. Birkhoff, G. (ed.): A Source Book in Classical Analysis. Harvard University Press, Cambridge (1973)
4. Cardano, G.: The Great Art (Ars Magna). Dover, New York (1993)
5. Cooke, R.: Classical Algebra: Its Nature, Origins, and Uses. Wiley, New Jersey (2008)
6. Ebbinghaus, H.-D., et al.: Numbers. Springer, Berlin (1990)
7. Katz, V.: A History of Mathematics, 3rd edn. Addison-Wesley, Boston (2009)
8. Katz, V., Parshall, K.H.: Taming the Unknown: A History of Algebra from Antiquity to the Early Twentieth Century. Princeton University Press, Princeton (2014)
9. Kline, M.: Mathematical Thought from Ancient to Modern Times Oxford University Press, Oxford (1972)
10. Mazur, B.: Imagining Numbers. Farrar, Straus and Giroux, New York (2003)
11. Nahin, P.G.: An Imaginary Tale: The Story of $\sqrt{-1}$. Princeton University Press, Princeton (1998)
12. Ore, O.: Cardano, The Gambling Scholar. Dover, New York (1965)
13. Struik, D.J. (ed.): A Source Book in Mathematics, 1200–1800. Harvard University Press, Cambridge (1969)
14. Tignol, J.-P.: Galois Theory of Algebraic Equations. World Scientific Publ., River Edge (2001)

Further Reading

15. Dobbs, D.E., Hanks, R.: A Modern Course on the Theory of Equations. Polygonal Publishing House, New Jersey (1980)
16. Pesic, P.: Abel's Proof: An Essay on the Sources and Meaning of Mathematical Unsolvability. MIT Press, Cambridge (2003)
17. Sesiano, J.: An introduction to the history of algebra: Solving equations from Mesopotamian times to the Renaissance. Am. Math. Soc. (2009)
18. Turnbull, H.W.: Theory of Equations. Oliver and Boyd, Edinburgh (1957)
19. van der Waerden, B.L.: Geometry and Algebra in Ancient Civilizations. Springer, Berlin (1983)
20. van der Waerden, B.L.: A History of Algebra, from al-Khwārizmī to Emmy Noether. Springer, Berlin 1985
21. Wussing, H.: The Genesis of the Abstract Group Concept. MIT Press, Cambridge (1984). (Translated from the German by A. Shenitzer)

i. The even natural numbers.
ii. The integers.
iii. The positive rational numbers (nontrivial for students).
iv. The rational numbers.
v. The algebraic numbers.

An "algebraic number" is a complex number which is a root of a polynomial equation with *integer* coefficients. Those complex numbers that are not algebraic are called "transcendental". Algebraic numbers are generalizations of rational numbers: the rational number m/n is a root of the equation $nx - m = 0$, m and n integers. The algebraic numbers are important especially in number theory (see ▶ Chapter 6).

To show that the algebraic numbers can be listed in a sequence, we associate with each polynomial $a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$, a_i integers, the rational number $2^{a_0} \times 3^{a_1} \times 5^{a_2} \times \ldots \times p_n^{a_n}$, where p_n is the $(n+1)$-st prime. This mapping is a 1–1 correspondence between all polynomials with integer coefficients and the positive rational numbers. Since each polynomial has finitely many roots, it follows that the algebraic numbers can be listed in a sequence.

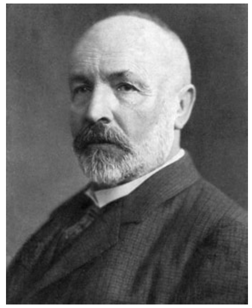
We have given various examples of sets which can be listed in a sequence. We refer to these sets as "denumerable" or "countable"—sets which can be enumerated. They all have the same cardinality, which we denote by $\aleph_0$ (aleph subzero; "aleph" is the first letter of the Hebrew alphabet). It is the smallest infinite cardinal. More generally, the cardinality of a set S is denoted by $|S|$; then, given two infinite sets S and T, $|S| \le |T|$ if there exists a 1–1 correspondence between S and a subset of T. If $|S| \le |T|$ but $|S| \ne |T|$, then $|S| < |T|$.

9.6 Paradoxes Regained

Let us now consider some *geometric* examples of infinite sets. We have seen that the points on two circles of unequal diameters have equal cardinality. It is therefore not surprising that any two line segments have the same cardinality. For example, the function $f(x) = 2x,\ x \in (a, b)$ gives a 1–1 correspondence between two intervals, one twice the other's length. What *is* surprising (shocking?) is that the points on a line segment, no matter how small, and on *the entire real line* have the same cardinality. The mapping $f(x) = \tan x$, $-\pi/2 < x < \pi/2$, gives a 1–1 correspondence between the interval $(-\pi/2, \pi/2)$ and the real line.

Another unexpected but fundamental result is that the real numbers are nondenumerable (uncountable), that is, their cardinality is greater than $\aleph_0$. It was only after fruitless attempts to prove the contrary that Cantor succeeded in establishing this. Here is a proof, though not Cantor's. Suppose that the real numbers in the interval (0, 1) *can* be written in a sequence, say a_1, a_2, a_3, …. Enclose each a_i in an interval of length $1/10^i$. Then the interval (0, 1) has been enclosed with intervals of total length $1/10 + 1/10^2 + 1/10^3 + \ldots = 1/9$, obviously a contradiction.

Yet another of Cantor's results (proved in the 1870s) which was contrary to prevailing opinion, and to "common sense", was that the real numbers and the complex numbers have the same cardinality. He found this "astonishing", given that the two sets are of different dimensions. He wrote to Richard Dedekind about it, exclaiming: "I see it but I don't believe it" [7, p. 126]. We prove here an equivalent result, namely that the line segment $A = (0, 1)$ and the square $B = (0, 1) \times (0, 1)$ have the same cardinality. Define mappings $f: A \to B$ by $f(0.r_1r_2r_3\ldots) = (0.r_1r_3r_5\ldots, 0.r_2r_4r_6\ldots)$ and $g: B \to A$ by $g\left(0.b_1b_2b_3, \ldots, 0.c_1c_2c_3\ldots\right) = \left(0.b_1c_1b_2c_2b_3c_3\ldots\right)$. With a little care to avoid duplication, these mappings establish a 1–1 correspondence between A and B.



Georg Cantor (1845–1918)

To give substance to his evolving ideas on the infinite, Cantor devised an arithmetic of completed infinities, a so-called *transfinite arithmetic*. The crucial idea, the cornerstone of his arithmetic, is the concept (used above, informally) of "1–1 correspondence", the comparison of sets "for size". If to each element of a set A there corresponds exactly one element of a set B, and conversely, to each element of B there corresponds a unique element of A, then a 1–1 correspondence between A and B is said to have been established, and the two sets are deemed to have the same number of elements, the same "cardinality".

9.4 Paradoxes Lost

This definition resolves Galileo's dilemma: the natural numbers and their squares do indeed have the same cardinality, for we can set up a 1–1 correspondence between them, as Galileo had done. The same goes for any two circles of unequal diameter: they too have the same cardinality. So these two examples do not give rise to paradoxes. Zeno's paradoxes are of a different sort, much more subtle than the others, claim some philosophers, who have been debating them, without resolution, for centuries [12].

What about the doctrine that "the whole is greater than any of its parts", employed implicitly by Galileo and the medieval scholars in arriving at their respective paradoxes? This doctrine is one of Euclid's common notions (axioms) given in his formulation of axiomatic geometry (see ▶ Chapter 2). It makes perfectly good sense for *finite* sets, but it must be abandoned for *infinite* sets. In fact Euclid's doctrine *never* holds for infinite sets: *every* infinite set contains a proper subset having the same cardinality as the original set. We can take this property to be the *definition* of infinite sets: a set S is "infinite" if it contains a proper subset with cardinality equal to that of S.

9.5 Denumerable (Countable) Infinity

Now to some examples. The elements of each of the sets below can be listed in a sequence (not necessarily by size), beginning with a first, second, third, and so on; hence the cardinality of each of these sets is the same as that of the natural numbers, 1, 2, 3, ….

Analytic Geometry: From the Marriage of Two Fields to the Birth of a Third

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics,* Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_3,

3.1 Introduction

Analytic geometry was invented independently by René Descartes and Pierre de Fermat in the first half of the seventeenth century. (The *term* "analytic geometry" was coined by Sylvestre François Lacroix in 1792.) The independent and more or less simultaneous invention (discovery) of concepts, results, or even major theories is not uncommon in mathematics; two other outstanding instances are calculus and noneuclidean geometry (see ▶ Chapters 5 and 7, respectively). As the mathematician Wolfgang Bolyai put it [1, p. 263]:

> » Mathematical discoveries, like springtime violets in the woods, have their season which no human can hasten or retard.

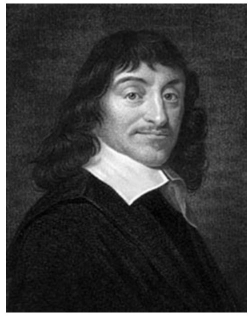
The analytic geometry of Fermat and Descartes did not originate in a vacuum. It drew inspiration from the Greek geometric tradition and, such as that was, the Renaissance algebraic tradition. The use of coordinates to designate position is possibly prehistoric, and appears in ancient Greek astronomy. Apollonius' great work *Conics* (ca. 250 BC) contains what are essentially the equations of these curves with respect to a fixed coordinate system, although not, of course, in modern notation. Furthermore, graphical representation of physical laws goes back to the scholar Nicole Oresme (ca. 1320–1382). What then was left for Descartes and Fermat to do? It was

(a) to recognize the following as a basic principle of analytic geometry: that to each geometric plane curve corresponds an equation in two unknowns, and, conversely, that to each equation involving two unknowns corresponds a curve; and
(b) to develop this principle into an algorithmic procedure, showing how it can be used to solve problems in geometry.

3.2 Descartes

Descartes wanted to devise a systematic method for the solution of geometric problems, especially those dealing with curves. Such a study received impetus from scientific developments in the early seventeenth century: Kepler's use of conic sections in his study of planetary motion, Galileo's use of parabolas to describe the motion of projectiles, and the use of curved lenses in the newly invented telescope and microscope.

René Descartes (1596–1650)



But the main motivation for the creation of analytic geometry came not from practical problems but from a desire to systematize the ancients' problem-solving tools. Descartes noted that many constructions and proofs in euclidean geometry called for new, inventive, and ad hoc approaches. He therefore undertook to exploit the power of algebra to provide a broad methodology for solving geometric problems.

Descartes was arguably the first great "modern" philosopher, as well as a first-rate mathematician and scientist. According to the distinguished historian of mathematics Henk Bos (1940–):

> » Descartes' mathematics was a philosopher's mathematics. From the earliest documented phase in his intellectual career, mathematics was a source of inspiration and an example for his philosophy, and, conversely, his philosophical concerns strongly influenced his style and program in mathematics [2, p. 228].

His great mathematical work—*Geometry (La Géométrie)*—appeared as one of the appendices to his philosophical treatise *Discourse on the Method of Reasoning Well and Seeking Truth in the Sciences*. In the latter work he sought a way to establish truths in all fields of endeavor. Geometry was identified as one of the three disciplines exhibiting that general method; the other two were meteorology and optics.

The essence of Descartes' method in geometry is given in several places in his book; here is one [3, p. 90]:

> » All points of a geometric curve [as defined by motions] must have a definite relation expressed by an equation.

In his analysis of geometric problems Descartes admitted only certain types of curves, namely those defined by "motions" or by loci [9, p. 483]. As an application of this method, he singled out for special attention the so-called Problem of Pappus: given four straight lines, to find the locus of a point that moves so that the product of its distances from two of the lines is in a fixed ratio to the product of its distances from the other two lines [11, p. 128]. Descartes

$$
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & \ldots. \\
\updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \cdots \\
1 & 4 & 9 & 16 & 25 & 36 & \ldots..
\end{array}
$$

Galileo concluded that the difficulties arise because

> » We attempt, with our finite minds to discuss the infinite, assigning to it those properties which we give to the finite and limited; but this … is wrong, for we cannot speak of infinite quantities as being the one greater or less than or equal to another. [12, p. 5]

Further rejection of the actual infinite came from Descartes, Spinoza, Leibniz, Hobbes, and Berkeley (see [1]). Even the great Gauss objected to its use, in a letter to his friend Schumacher in 1831:

> » I protest against the use of an infinite quantity as an actual entity; this is never allowed in mathematics. The infinite is only a manner of speaking, in which one properly speaks of limits to which certain ratios can come as near as desired, while others are permitted to increase without bound. [8, p. 160]

9.3 Cantor

Modern understanding of the mathematical infinite is the almost singlehanded creation of Georg Cantor. Cantor urged that the old distinction between the potential and the actual infinite is dubious: "in truth the potentially infinite concept has only a borrowed reality, insofar as a potentially infinite concept always points toward a logically prior actually infinite concept whose existence it depends on" [12, p. 3]. His revolutionary approach stems from 1870, when, on the urging of a colleague at the University of Halle, he started doing research on trigonometric series, following a PhD in number theory. Two years later, at the age of 27, he wrote a paper on the subject, in particular on the question of *unique* representation of functions in such series. He found that in this research he needed a proper understanding of the real numbers, which was then lacking. The result was his now well known representation of the reals as Cauchy (fundamental) sequences. The latter entailed an encounter with the actual infinite, for a Cauchy sequence is an infinite collection of rational numbers satisfying given conditions. While previously opposed to the notion of a completed infinity (as was everyone else), Cantor soon realized that he could make little progress in his researches without accepting it. He set aside his work on trigonometric series to devote all his time to the development of what is now known as *transfinite set theory*. Here are some of his thoughts on the matter [4, p. 211]:

> » It is traditional to regard the infinite as the indefinitely growing or in the closely related form of a convergent sequence, which it acquired during the seventeenth century. As against this I conceive the infinite in the definite form of something consummated, something capable not only of mathematical formulation, but of definition by number. This conception of the infinite is opposed to traditions which have grown dear to me, and it is much against my own will that I have been forced to accept this view. But many years of scientific speculation and trial point to these conclusions as to a logical necessity, and for this reason I am confident that no valid objection could be raised which I would not be in position to meet.

Aristotle (384–322 BC)

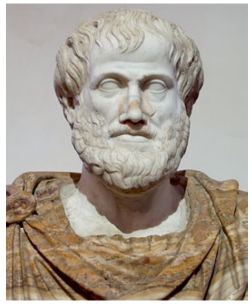
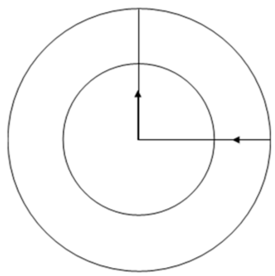


Figure 9.1 Two concentric circles with unequal diameters but with equal numbers of points

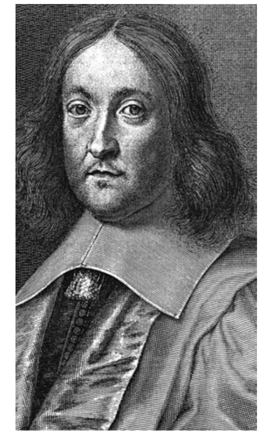


9.2 Before Cantor

In the Middle Ages the few discussions of infinity appeared mostly in theological contexts (cf. our expressions "God almighty" and "God, in his infinite wisdom"). There was however a *mathematical* example in the thirteenth century, after scholars noted that two circles of unequal diameter have an equal number of points but (clearly) unequal perimeters. This was evidently a paradox. The former observation is ascertained by establishing a one-to-one correspondence which matches each point of one circle with exactly one point of the other: place the circles so that they are concentric, and then the 1-1 correspondence is established by having the center of the two circles play the crucial role [6]. (See Figure 9.1).

Our next encounter with the actual infinite comes in the seventeenth century, at the dawn of the modern period in the history of mathematics. In his book *Dialogues Concerning Two New Sciences* (1638), Galileo Galilei pondered the contradictions that arise when one tries to compare (for "size") the set of positive integers and the set of their squares. On the one hand, he argued, there are clearly more of the former than of the latter; on the other, one can set up a 1–1 correspondence between the two sets of numbers which matches each element in the first with a unique element in the second, as follows:

Pierre de Fermat (1601-1665)



showed that the locus is a conic section [7, p. 87]. This result was already known to the Greeks [2, ► Chapter 23]. To exhibit the substantial power of his method, Descartes generalized the Problem of Pappus to the case of 2n lines and derived the equation of the locus for small values of n; for $n = 3$, he showed that the equation is of degree 3. But he showed little interest in the shapes of the curves given by such equations. While the salient idea for the subsequent development of mathematics was the association of equation and curve, for Descartes the idea was just a means to an end—the solution of geometric problems.

3.3 Fermat

By the beginning of the seventeenth century the extant Greek mathematical works had been restored and had elicited great interest. Fermat introduced his new method in geometry after a careful study of the geometric works of Apollonius (ca. 225 BC) and Pappus (ca. 300 AD) and of the algebraic work of Viète. He noted that although the Greeks studied loci, they must have found them difficult, since some of the problems were not stated in full generality. He proceeded to rectify this in a twenty-page work titled *Introduction to Plane and Solid Loci*. The basic principle of analytic geometry is stated at the outset [3, p. 75]:

> Whenever in a final equation two unknown quantities are found, we have a locus, the extremity of one of these describing a line, straight or curved.

The historian of mathematics Carl Boyer (1906–1976) referred to this sentence as "one of the most significant statements in the history of mathematics", for it introduced "not only analytic geometry, but also the immensely useful idea of an algebraic variable" [3, p. 75]. Fermat's work had considerable influence on, among others, Newton and Leibniz.

The work of Fermat and Descartes had different emphases. While Descartes stressed the fact that curves can be represented by equations, Fermat's point of departure was that indeterminate equations give rise to curves. He showed that equations of the first degree with two variables describe straight lines, and he carefully analyzed equations of the second degree with two variables, showing that they represent various conic sections. Although he did not consider equations of degree higher than two, he clearly recognized the potential of the subject he was dealing with to produce new curves, as is evident from his statement that "the species of curves are indefinite in number: circle, parabola, hyperbola, ellipse, etc." [3, p. 79].

3.4 Descartes' and Fermat's Works from a Modern Perspective

Although the analytic geometry of Descartes and Fermat was groundbreaking, it was not in the form now familiar to us. In particular:

(a) Remarkably, a rectangular coordinate system and formulas for distance and slope are missing. In fact, coordinate axes are not explicitly set forth. Only the horizontal axis appears explicitly in drawings, while the implicit vertical axis is usually oblique.

(b) The unknowns x and y which appear in the equation of a curve were considered to be line segments rather than numbers. It was not until a century or more later that coordinates began to be viewed as numbers. The notion of a one–one correspondence between points in a plane and ordered pairs of real numbers, nowadays the basis of our formulation of analytic geometry, was foreign to Fermat and Descartes.

(c) Descartes considered only curves whose equations are "algebraic" (that is, polynomials in x and y). Transcendental curves, such as $y = \log x$, $y = \sin x$, and $y = e^x$, did not come under the scope of his general method. Fermat, as we noted, confined himself essentially to polynomial equations of degree two in x and y. (In another work, Fermat also considered the so-called higher parabolas and hyperbolas, $y = x^n$ and $y = x^{-n}$, respectively.)

(d) Curve-sketching in the sense familiar to us was not a central aspect of the analytic geometry of Fermat and Descartes. Fermat emphasized the study of equations in x and y not via their graphical representation but via their properties as derived by the methods of calculus. Descartes (we recall) did not regard the equation of a curve as an adequate *definition* of the curve.

(e) Both Descartes and Fermat used only positive coordinates, and such curves as were sketched appeared only in the first quadrant. Negative numbers were not a commonly acceptable part of the number system. Moreover, since Descartes' objective, and to a large extent Fermat's, was to solve geometric problems, the need for negative coordinates did not arise.

At first the geometry of Descartes and Fermat was accessible only to a very small circle of the ablest mathematicians. The latter did not take kindly and quickly to the idea of algebra, conceived as a collection of formulas and rules of manipulation, playing the dominant role in the rigorous, axiomatic, venerable field of geometry. It is only with Gaspard Monge and Lacroix in the latter part of the eighteenth century that we find analytic geometry essentially as it appears in today's textbooks. In the intervening years, analytic geometry was developed by, among others, Leibniz, who introduced transcendental curves into the study of geometry; Newton, who used negative coordinates freely, sketched curves from their equations, and introduced various

The Infinite: From Potential to Actual

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics*, Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_9,

David Hilbert, arguably the greatest mathematician of the first half of the twentieth century, voiced memorably the fascination and challenge of the mathematical infinite [11, p. vii]:

> » The infinite! No other question has ever moved so profoundly the spirit of man; no other idea has so fruitfully stimulated his intellect; yet no other concept stands in greater need of clarification than that of the infinite.

9.1 The Greeks

The challenges loomed early in the history of Greek mathematics. Consider the attempt by Democritus to calculate the volume of a cone by regarding it as composed of thin slices parallel to its base. If the number of slices is finite, and the thickness of each slice is nonzero, then the surface of the cone will appear "stepped", not smooth—which implies that the true volume is somehow a sum of infinitely many zeroes (see for example [3], pp. 79–81). The four famous paradoxes of Zeno (ca 450 BC), which probably aimed to support the claim of his teacher Parmenides that motion is impossible, are no less perplexing. In the "Dichotomy", for example, Zeno argues that to move from point A to point B, one must first get halfway to B, then halfway to the remaining distance, and so on. Assuming that space, and in particular the line segment AB, is infinitely divisible, it follows that one must cover infinitely many steps in finite time. But, Zeno claims, this is clearly impossible—so motion is impossible.

Among the ancient Greek thinkers it was Aristotle who considered the infinite most deeply. He concluded that any geometrical magnitude, such as a line segment, is infinitely divisible, for (he said) the idea of a minimum magnitude makes no sense. Similarly, the set of numbers—which for Aristotle included only the *natural* numbers 1, 2, 3, …—can clearly be extended as far as we please. Time has both of these properties: it extends without limit, and any portion of it can be divided without limit. These Aristotelian views were shaped by considerations outside mathematics, for example the great philosopher's belief that time can have neither a beginning nor an end.

But Aristotle went further, to a fundamental distinction of great importance. He held, for example, that although we can push the set of natural numbers arbitrarily far, we cannot grasp their totality as a single entity. This difference between the "potential" infinite and the "actual" infinite appears also in geometry, where, Aristotle urged, a straight line cannot be infinite but a mathematician can extend it as far as he/she needs or pleases. This avoidance of "actual" infinities undoubtedly reflects the close tie of Greek thought to the physical world—where of course we do not experience the infinite. See [1].

Further Reading

13. Bashmakova, I.G., Smirnova, G.S.: The Beginnings and Evolution of Algebra. The Mathematical Association of America, Washington, D.C. (2000) (Translated from the Russian by A. Shenitzer.)
14. Berlinghoff , W.P., Gouvêa, F.Q.: Math Through the Ages: A Gentle History for Teachers and Others, Expanded Edition. The Mathematical Association of America: Oxton House Publishing, Washington, D.C. (2004)
15. Dobbs, D.E., Hanks, R.: A Modern Course on the Theory of Equations. Polygonal Publishing House, Passaic (1980)
16. Katz, V.: A History of Mathematics: An Introduction, 3rd edn. Addison-Wesley, Boston (2009)
17. Mazur, B.: Imagining Numbers. Farrar, Straus, and Giroux, New York (2003)
18. Turnbull, W.: Theory of Equations. Oliver & Boyd, New York (1957)
19. van der Waerden B.L.: A History of Algebra. Springer-Verlag, Berlin (1985)

Leonhard Euler (1707–1783)

coordinate systems, among them the polar; and Euler, who developed three-dimensional analytic geometry (already hinted at by Descartes and Fermat), and who did much to systematize the subject in his outstanding book of 1748, *Introductio in Analysin Infinitorum*.

3.5 The Significance of Analytic Geometry

Descartes' and Fermat's founding of the subject was revolutionary, although initially it might not have been viewed as such. Several decades after its creation, the new subject/method laid the mathematical groundwork for calculus and Newtonian physics. More specifically:

(a) Fermat and Descartes were the first to highlight the very important notion of a (continuous) variable—indispensable in the development of calculus.
(b) The use of equations to define curves opened up the possibility of introducing an unlimited number of new curves, beyond the conception of the synthetic method. Such curves, in turn, called for the invention of algorithmic techniques for their systematic investigation—an important factor in the creation of calculus.
(c) Undoubtedly the study of the physical world calls for geometric knowledge: objects in space are geometric figures, paths of moving bodies are curves. Analytic geometry made possible the expression of shapes and paths in algebraic form, from which quantitative knowledge can be derived.

Analytic geometry produced a most important coupling of algebra and geometry—a relationship that proved very fruitful for subsequent developments in mathematics. Lagrange expressed it as follows [11, p. 322]:

> » As long as algebra and geometry travelled separate paths, their advance was slow and their application limited. But when these two sciences joined company, they drew from each other fresh vitality and thenceforward marched at a rapid pace toward perfection.

The mathematician Keith Kendig (1938–) echoed these remarks [10, p. 161]:

» [Analytic geometry] gave our imagination 'two ends'—an algebraic one and a geometric one; geometric insight could often be translated into an algebraic one, and vice versa.

Morris Hirsch (1933–), another prominent mathematician, was more specific [8, p. 604]:

3

» If geometry lets us see what we are thinking about, algebra enables us to talk precisely about what we see, and above all to calculate. Moreover, it tends to organize our calculations and to conceptualize them; this, in turn, can lead to further geometrical construction and algebraic calculation.

Linear algebra is another excellent example of the interplay of algebra and geometry. For instance, the algebraic formulation of dimension makes natural the extension to dimensions higher than three. On the other hand, speaking about "lines" and "planes" in dimensions higher than three makes the subject more intuitive, suggestive, and comprehensible.

Analytic geometry—a bridge between algebra and geometry—also provides bridges between shape and quantity, number and form, the analytic and the synthetic, the discrete and the continuous. For, as was shown in the nineteenth century, the real numbers can be built up rigorously from the integers, and since the one–one correspondence between the real numbers and the points on a line is at the root of analytic geometry, this establishes a bridge between the continuous and the discrete. This correspondence—this tension—has been most fruitful in the development of mathematics. Hermann Weyl, one of the foremost mathematicians of the first half of the twentieth century, noted that it "represents a remarkable link between something which is given by our spatial intuition and something that is constructed in a purely logico-conceptual way" [5, p. 159].

In the twentieth century such bridge-building became enormously important, offering powerful tools to mathematicians. As examples, consider the following disciplines, which by merging two fields lent strength to each: analytic number theory, differential topology, geometric number theory, algebraic topology, algebraic number theory, differential geometry, and algebraic geometry. A grand synthesis—the Langlands Program—relating several areas of mathematics, in particular number theory, algebra, and analysis, was proposed by Robert Langlands (1936–) in the 1960s in a series of deep and far-reaching conjectures, some by now established [6].

Problems and Projects

1. Discuss the thesis, advocated by some historians, that the Greeks invented analytic geometry. Consult [2, 3, 9, 11].
2. What is the origin of the words "ellipse", "hyperbola", and "parabola"? See [2, 3, 9, 11].
3. How did Descartes solve the four-line locus problem of Pappus? See [2, 7, 9].
4. Discuss the coordinate systems (such as they were) of Descartes and Fermat. See [2, 3, 4, 9, 11, 12].
5. Descartes' geometry contains much on the theory of equations, especially in the third of the book's three chapters. Describe it. See [2, 7, 9, 11].
6. Write a brief biography of either Descartes or Fermat.
7. Describe the principles, as outlined in Descartes' major work in philosophy, *Discourse on Method*, which were based on his general method of acquiring knowledge. See [2, 3, 7, 9].
8. Discuss some contributions to analytic geometry of the successors of Fermat and Descartes. See [2, 3, 9, 11].

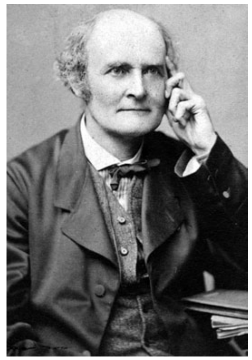
2. (i) Define the product of quaternions, represented as quadruples $a+bi+cj+dk$, and show that every nonzero quaternion has an inverse.
   (ii) Show that **K** (the octonions) are not associative.
3. (i) Students with some background in abstract algebra may find it interesting to show that the only n-tuples of reals that form *associative* division algebras are the real numbers, the complex numbers, and the quaternions (see [3, 6]).
   (ii) There is an "elementary" proof which shows that, for *odd* n, a division algebra of n-tuples of reals is possible only for $n=1$ (see [3, p. 190]).
4. (i) There is an important product defined on triples of reals, namely the *vector product*: $(a_1i+a_2j+a_3k)\times(b_1i+b_2j+b_3k)=(a_2b_3-a_3b_2)i+(a_3b_1-a_1b_3)j+(a_1b_2-a_2b_1)k$. Show that the vector product (×), the quaternion product (*), and the scalar (inner) product (·) of 3-dimensional vectors are related: $\alpha\times\beta=\alpha*\beta+\alpha\cdot\beta$ [2, 6]. The only other Euclidean n-space in which a "cross product" can be defined is the space with $n=7$ [7].
   (ii) The historian of mathematics Michael Crowe argues that the quaternions were instrumental in the creation of vector analysis. Vector analysts and quaternionists were at loggerheads during the second half of the nineteenth century about the preferable way to deal with problems in physics. Write an essay discussing this issue (see [2, 6]).
5. Defining "integral quaternions" and using ideas from number theory (unique factorization), one can prove Lagrange's theorem that every positive integer is a sum of four squares (of integers). Outline the ideas involved in such a proof (see [3, 5]).
6. Write a brief account of the life and work of Hamilton (see [2, 3, 4]).
7. The quaternions and the octonions are (hypercomplex) "numbers" – and of course the integers, rationals, reals, and complex numbers are numbers. Are the polynomials (over the reals, say) numbers? The integers modulo m? What might (some of) these "numbers" have in common? What is a "number", anyway? Research this topic.

8

References

1. Cooke, R.: Classical Algebra: Its Nature, Origins, and Uses. Wiley, Hoboken (2008)
2. Crowe, M.J.: A History of Vector Analysis. University of Notre Dame Press, Notre Dame (1967)
3. Ebbinghaus, H.-D.: Numbers. Springer-Verlag, New York (1990)
4. Hankins, T.L.: Sir William Rowan Hamilton. The Johns Hopkins University Press, Baltimore (1980)
5. Herstein, I.N.: Topics in Algebra. Blaisdell Publishing Co., New York (1964)
6. Kantor, I.L., Solodovnikov, A.S.: Hypercomplex Numbers: An Elementary Introduction to Algebras. Springer-Verlag, New York (1989) (Translated from the Russian by A. Shenitzer.)
7. Massey, W.: Cross products of vectors in higher dimensional euclidean spaces. Am. Math. Mon. **90,** 697–701 (1983)
8. Niven, I.: Equations in quaternions. Am. Math. Mon. **48,** 654–661 (1941)
9. Niven, I.: The roots of a quaternion. Am. Math. Mon. **49,** 386–388 (1942)
10. Poincaré, H.: Poincaré's review of Hilbert's foundations of geometry. Bull. Am. Math. Soc. **37,** 77–78 (1999) (Reprinted)
11. van der Waerden B.L.: Hamilton's discovery of quaternions. Math. Mag. **49,** 227–234 (1976)
12. Wussing, H.: The Genesis of the Abstract Group Concept. MIT Press, Cambridge (1984) (Translated from the German by A. Shenitzer.)

Arthur Cayley (1821–1895)



It is tempting to continue in this manner and define an algebra of 16-tuples in the hope that it will turn out to be a division algebra; but any such attempt is doomed to failure. We have the following sequence of four theorems which address these issues; the first two were proved in the late nineteenth century and the other two in the twentieth (see [3, 6]).

i. The only n-tuples of reals that form associative and commutative division algebras are **R** and **C** (that is, those n-tuples for which $n = 1$ or 2), where **R** and **C** denote the real and complex numbers, respectively.
ii. The only n-tuples of reals that form associative (but not necessarily commutative) division algebras are **R**, **C**, and **H** (that is, $n = 1, 2, 4$).
iii. The only n-tuples of reals that form alternative (but not necessarily associative or commutative) division algebras are **R**, **C**, **H**, and **K** (that is, $n = 1, 2, 4, 8$).
iv. The only n-tuples of reals that form division algebras (not necessarily alternative, associative, or commutative) are those for which $n = 1, 2, 4, 8$. (Such algebras need not be any of **R**, **C**, **H**, or **K**, but their *dimensions* must be one of 1, 2, 4, or 8.) The proof of this result uses topology.

Other important "algebras" (number systems) were defined in the decades following the introduction of the quaternions—for example, exterior algebras (Hermann Grassmann), group algebras (Cayley), matrices (Cayley), triple algebras (Augustus De Morgan), and biquaternions (Clifford). In time, and motivated by some of these examples, the general concept of a "hypercomplex number system"—an *associative algebra*—emerged, and became one of the pillars of a newly established field—"abstract algebra" (see [3, 6]).

Problems and Projects

1. The Fundamental Theorem of Algebra says that every polynomial with complex coefficients has a complex root. An analogous result holds with "complex" replaced by "quaternion". Note however that a polynomial of degree n over the quaternions need not have n quaternion roots. For example, $x^2 + 1 = 0$ has infinitely many roots: $bi + (1-b^2)^{1/2}j$, where b is a real number with $0 \le b \le 1$. Write a report on the issue of roots of polynomial equations with *quaternion* coefficients (see [3, 8, 9]).

References

1. Bell, E.T.: The Development of Mathematics, 2nd edn. McGraw-Hill, New York (1945)
2. Bos, H.J.M.: Redefining Geometrical Exactness: Descartes' Transformation of the Early Modern Concept of Construction. Springer, New York (2001)
3. Boyer, C.B.: History of Analytic Geometry. Scripta Mathematica, New York (1956)
4. Eves, H.: Great Moments in Mathematics (Before 1650). Mathematical Association of America, Washington DC (1980)
5. Gardiner, A.: Infinite Processes: Background to Analysis. Springer-Verlag, Berlin (1982)
6. Gelbart, S.: An elementary introduction to the Langlands Program. Bull. Am. Math. Soc. **10**, 177–219 (1984)
7. Grabiner, J.: Descartes and problem solving. Math. Mag. **68**, 83–97 (1995)
8. Hirsch, M.: Review of Linear Algebra Through Geometry, by T. Banchoff and J. Wermer. Am. Math. Mon. **92**, 603–605 (1985)
9. Katz, V.: A History of Mathematics: An Introduction, 3rd edn. Addison-Wesley, Boston (2009)
10. Kendig, K.M.: Algebra, geometry, and algebraic geometry. Am. Math. Mon. **90**, 161–174 (1983)
11. Kline, M.: Mathematical Thought from Ancient to Modern Times. Oxford University Press, Oxford (1972)
12. Mahoney, M.: The Mathematical Career of Pierre de Fermat, 2nd edn. Princeton University Press, Princeton (1994)

Further Reading

13. Descartes, R.: The Geometry of René Descartes, with a Facsimile of the First Edition. Dover, New York (1954)
14. Scott, J.F.: The Scientific Work of René Descartes (1596–1650). Taylor & Francis, New York (1952)

Probability: From Games of Chance to an Abstract Theory

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics*, Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_4, © Springer Science+Business Media, LLC 2015

4.1 The Pascal–Fermat Correspondence

Probability, like various other mathematical concepts and theories, emerged from the desire to solve real-world problems—in this case, to provide a mathematical framework for games of chance and for gambling. One must of course distinguish between "probability" as a concept and "probability" as a subject. We have occasionally used "probability theory" for the latter term. Normally "probability" is used for both concept and theory, the context making clear which is intended.

Gambling is a long-standing activity, going back over three thousand years and engaged in by all civilizations. But the *mathematical analysis* of gambling, leading to the advent of probability, is of relatively recent origin. It began with Blaise Pascal and Pierre de Fermat. Around 1653 Pascal was approached by Antoine Gombauld, Chevalier de Méré (1607–1684)—a man of letters with a considerable knowledge of mathematics—to help him solve two gaming problems. Many writers say the request was made to improve de Méré's gambling chances [3, p. 84]; Oysten Ore disputes that claim [14].

The two problems came to be known as the "Dice Problem" and the "Division Problem", the latter also known as the "Problem of Points".

The Dice Problem: How many throws of two dice are needed in order to have a better-than-even chance of getting two sixes?

The Division Problem: What is a fair distribution of stakes in a game interrupted before its conclusion?

The Dice Problem is much the simpler of the two. It was solved in the mid-sixteenth century by Girolamo Cardano, among others, using plausibility arguments. Cardano, called "The Gambling Scholar" by Ore [13], was a colorful figure. A physician and mathematician by profession, he was also a practicing astrologer and an inveterate gambler, who composed "a learned book on games and ways to win in gambling" [13, p. viii], entitled *The Book on Games of Chance* [13]. His most influential work, *The Great Art,* dealing with the solution of the cubic and quartic equations by radicals, was a fundamental contribution to mathematics (see ▶ Chapter 2).

The Division Problem presents a much greater challenge. Among the first to introduce it was the Italian mathematician Luca Pacioli, in his 1494 book *Everything About Arithmetic, Geometry, and Proportions,* better known as *Suma*. Here is his version of the problem [10, p. 489]:

> » Two players are playing a fair game [the players are equally capable] that was to continue until one player had won six rounds. The game stops when the first player has won five rounds and the second player three. How should the stakes be divided between the two players?

William Rowan Hamilton (1805–1865)

both events were radical departures from existing conceptions, and both led to fundamental developments in their respective fields.

8.4 Beyond the Quaternions

Like all revolutions, this one was not universally acclaimed. For example, John Graves, Hamilton's mathematician friend, said of the quaternions: "I have not yet any clear views as to the extent to which we are at liberty arbitrarily to create imaginaries [referring presumably to the i, j, k], and to endow them with supernatural properties" [2, p. 34]. But most mathematicians, including Graves, quickly came around to Hamilton's point of view. His quaternions served as a catalyst for the exploration of diverse "number systems" with properties which differed in various ways from those of the real and complex numbers.

First and foremost among such systems were the "octonions" or "Cayley numbers", discovered independently by Graves and by Arthur Cayley very soon after the discovery of the quaternions. These are 8-tuples of reals, containing the quaternions, which form a division algebra **K**. It is instructive to view **K** in the following manner:

Note first that the quaternions can be viewed as pairs of complex numbers: $a+bi+cj+dk = (a+bi)+(c+di)j = w+zj$, where $w, z \in \mathbf{C}$, $j^2=-1$. Now define multiplication of these pairs: $(w_1+z_1j)\ (w_2+z_2j) = (w_1w_2 - z_2{}^*z_1) + (z_1w_2{}^* + z_2w_1)j$, where z^* denotes the conjugate of z. (It is important to have the w_i and z_i above in precisely this order.) Verify that the product thus defined is the same as the usual product of quaternions given in terms of i, j, and k. These pairs of complex numbers are therefore the elements of **H**.

Let now $\mathbf{K} = \{\alpha+\beta e\colon \alpha, \beta \in \mathbf{H}\}$, where e is an arbitrary unit with $e^2=-1$. Define a product in **K** as follows: $(\alpha_1+\beta_1e)(\alpha_2+\beta_2e) = (\alpha_1\alpha_2 - \beta_2{}^*\beta_1) + (\beta_1\alpha_2{}^* + \beta_2\alpha_1)e$ (see the definition above of the product in **H**; the conjugate α^* of the quaternion $\alpha = a+bi+cj+dk$ is $a-bi-cj-dk$). These are the "octonions". They can be viewed of course as 8-tuples of reals. Since **K** contains **H**, it is clearly noncommutative. But it is also *nonassociative*, that is, there are a, b, c in **K** for which $a(bc) \neq (ab)c$; for example, $(ij)e \neq i(je)$. **K**, however, is "alternative", that is, $(xy)y = x(yy)$ and $y(yx) = (yy)x$ for all x, y in **K** (see [6]).

8.3 The Quaternions

Having defined the complex numbers as ordered pairs of reals, it was natural for Hamilton to inquire whether an algebra of triples exists to represent vectors in 3-space. Since the complex numbers were fundamental in many branches of mathematics and its applications, he considered the search for such an algebra of triples, which he had begun to pursue already in 1828, to be of vital importance.

Addition and subtraction of triples were to be defined componentwise in the obvious way. As for multiplication, Hamilton imposed several conditions it would have to satisfy: it had to be associative, commutative, and distributive (over addition); division had to be possible; the "law of the moduli" had to hold (the modulus of the product equals the product of the moduli, where the modulus of the triple (a, b, c) is $a^2+b^2+c^2$); and, finally, the product of triples had to have geometric significance, just as the product of vectors in the plane does [2].

Hamilton tried for fifteen years to define a multiplication on triples which would satisfy the conditions stated above; for a blow-by-blow account of his struggles, see [11]. As we know, he did not succeed, and turned instead to quadruples (a, b, c, d) of reals, which he also denoted, as an aid to computation, by a + bi+ cj + dk. These did indeed satisfy all the requirements he demanded of triples, *except for commutativity under multiplication*. Addition and subtraction were defined componentwise, the associative and distributive laws for quadruples were assumed, and the symbols i, j, and k were to satisfy the relations $i^2=j^2=k^2=ijk=-1$—these relations turned out to be necessary if multiplication was to "work" as Hamilton required [11]. Using the associative and distributive laws, the product of i, j, and k can be extended to all quadruples a + bi + cj + dk.

Hamilton called these objects "quaternions". We denote them by **H**. They form a "skew field"—that is, they satisfy all the axioms of a field except for commutativity of multiplication. In fact, they form a "division algebra"—a skew field that is also a vector space (over the reals, in this case). From the above identities follow the identities $ij=k=-ji$, $jk=i=-kj$, and $ki=j=-ik$. In particular, $ij \neq ji$, so that, indeed, the quaternions do not commute.

It is interesting to observe that had Hamilton suspected that triples a + bi + cj would not work (that is, would not yield a skew field extending the complex numbers), he could easily have *proved* that, and saved himself fifteen years of labor! Here is a proof: suppose that such a multiplication of triples is possible. Let ij = a + bi + cj, for some real numbers a, b, and c. Then i(ij) = i(a + bi + cj). Multiplying and collecting terms yields $c^2 + 1 = 0$—a contradiction. The benefits of hindsight!

In 1843 Hamilton presented his work on quaternions to the Royal Irish Academy. For the next twenty-two years, he was preoccupied almost exclusively with their application, mainly to geometry and physics. To him they were the long-sought key which would unlock the mysteries of these subjects. "I still must assert", he noted in 1851, "that this discovery appears to me to be as important for the middle of the nineteenth century as the discovery of fluxions [calculus] was for the close of the seventeenth" [2, p. 30].

But from a more modern perspective the importance of the quaternions lies in *algebra*. Their invention (discovery?) was a breakthrough in that subject's evolution. It detached the laws of algebra from those of arithmetic, the laws which the real and complex numbers obey. Now there was a "number system" which satisfied all the laws of arithmetic save for commutativity of multiplication. Henri Poincaré referred to this development as "a revolution in arithmetic quite comparable to that which Lobachevsky effected in geometry" [10, p. 78]. Indeed,

Blaise Pascal (1623–1662)

Pacioli claimed that the stakes should be split in the ratio of 5:3, which is incorrect. Unsuccessful attempts to solve the problem were also made by several Italian mathematicians of the sixteenth century, including Cardano and Niccolò Tartaglia. Tartaglia ventured that the stakes should be divided in the ratio 2:1.

And so we come to Pascal and the Chevalier de Méré. Pascal was intrigued by the Problem of Points proposed by de Méré and agreed to study it. Before long he had a solution. The problem was challenging and subtle, and so he wrote (in July 1654) to Fermat, the leading mathematician of France, asking if he would read his (Pascal's) solution. Fermat obliged. Thus began the now famous Pascal–Fermat correspondence, lasting several months (July–November 1654), and resulting in the emergence of what turned out to be a most important mathematical discipline—*probability*. Seven letters of that correspondence are extant, though it is not known how many were exchanged; in particular, Pascal's initial letter to Fermat is lost. The renowned mathematician and master probabilist Alfréd Rényi reconstructed four of Pascal's letters to Fermat [15].

What did the letters in the Pascal–Fermat correspondence contain? First, what they did *not* contain: there are no formal definitions, nor proofs of theorems; even the word "probability" does not appear (it will first show up about a century later). What we have are solutions of a problem—the Division Problem (the Problem of Points), its specializations and extensions (for example, to three players), different ways of looking at it, give-and-take between two brilliant mathematicians, the emergence of approaches to the solution of the problem, some combinatorics, and crucial ideas such as a fair die, equally likely events, and "favorable" events—although these last two important ideas are already present in the solutions of gaming problems by Cardano and by Galileo. The thrust of the Pascal–Fermat correspondence was that it put in motion what came to be known as "probability theory"—a new branch of mathematics. Indeed, contemporary mathematicians recognized that Fermat and Pascal had done precisely that.

Why was there no definition of probability, and why were there no theorems and proofs in the Pascal–Fermat correspondence—a work that initiated a new subject? First, such matters are not to be expected in a correspondence dealing with the solution of problems. Beyond that, it is almost always the case that the formal development of a subject comes at the *end* of an evolutionary process. Calculus is an excellent example of this phenomenon (see ▸ Chapter 5).

Pierre de Fermat (1601–1665)

Probability theory followed a similar route. Mathematicians knew very well what "probability" meant without having to define the concept, and they put the ideas of probability theory to excellent use in the eighteenth and nineteenth centuries without having a formal structure of the subject, which was introduced in the early twentieth century.

4.2 Huygens: The First Book on Probability

Christiaan Huygens was a first-rate Dutch mathematician, physicist, astronomer, and inventor. (Most mathematicians in the seventeenth and eighteenth centuries were also scientists.) He studied mathematics and law at the University of Leiden. On a visit to Paris in 1655 he became acquainted with the Problem of Points, though not with its solution. Taken with the problem, he promptly solved it. But he realized that one was dealing here with important ideas beyond the solution of problems. So he decided to write a book which would give expression to this broader point of view [9, p. 65]:

> I would like to believe that if someone studies these things a little more closely, then he will almost certainly come to the conclusion that it is not just a game which has been treated here, but that the principles and the foundations are laid of a very nice and very deep speculation.

The result of these speculations was a sixteen-page treatise titled *On Reckoning at Games of Chance,* published in 1657. Here is how a historian of the subject, Florence Nightingale David (1909–1993), saw this work [3, p. 110] (but see [5, p. 138 ff.] for a contrary view):

at about the same time by William Rowan Hamilton, was whether one can enlarge the number system beyond the complex numbers [4]. Both problems were among a select few that gave rise in the late decades of the nineteenth century and the early decades of the twentieth century to what has come to be known as "abstract algebra" [1]. We will consider in this chapter the second question: are there numbers beyond the complex numbers? To provide a context we must first discuss Hamilton's work on complex numbers.

8.2 Hamilton and Complex Numbers

The complex numbers were conceived by the Renaissance mathematician Rafael Bombelli and expounded in his book *Algebra* of 1572. The motivation for their introduction was the desire to solve polynomial equations, in particular the cubic. It took another two and a half centuries, and the imprimatur of Gauss, who in 1831 gave their geometric representation as points (or vectors) in the plane, to have them accepted as bona fide mathematical entities. Similar geometric representations of complex numbers were given by Caspar Wessel in 1797 and by Jean Robert Argand in 1806, among others, but their work went largely unnoticed (see ► Chapter 2).

Hamilton was the greatest Irish mathematician. He was a precocious child, who at the age of thirteen knew (besides English) thirteen languages: Greek, Latin, Hebrew, Syriac, Persian, Arabic, Sanskrit, Hindustani, Malay, French, Italian, Spanish, and German. Aside from languages, he studied geography, religion, literature, astronomy, and mathematics. He read Euclid in Greek, Newton in Latin, and Laplace in French. At seventeen he found an error in the latter's renowned *Mécanique Céleste*.

Hamilton made outstanding contributions in optics, dynamics, and algebra. His interest in algebra was aroused around 1826 by his mathematician friend John Graves. Hamilton was dissatisfied with the *geometric* representation of complex numbers given by Gauss and others. After all, he observed, these are *numbers,* which he believed to be the domain of *algebra*. He objected in particular to the dependence of a geometric representation of complex numbers on a coordinate system. He was also unhappy with their representation as expressions of the form $a + bi$ (which Gauss, among others, had used). It seemed to him that adding bi to a was like adding oranges to apples. And what in any case is i, he asked?

These misgivings prompted Hamilton to define (in 1837) complex numbers as ordered pairs of reals. He defined, in the way that we still do, the four algebraic operations on pairs, and showed that under these operations the ordered number-couples come close to satisfying the laws of what we now call a field: they obey the closure laws and the commutative and distributive laws (he introduced the associative law a decade later in his work on quaternions); moreover, these pairs possess additive and multiplicative inverses, and they include a zero element.

This was a substantial conceptual advancement in algebra, given that in the mid-1820s the subject consisted largely of rules for the manipulation of algebraic expressions, especially those involving negative and complex numbers, and the solution of polynomial equations. Note, for example, that in Hamilton's version of complex numbers the "mysterious" i is just the "ordinary" pair (0, 1).

Hypercomplex Numbers: From Algebra to Algebras

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics*, Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_8, © Springer Science+Business Media, LLC 2015

8.1 Introduction

The *term* "algebra" dates back to the ninth century AD, but the *subject*, referring to the solution of polynomial equations, is roughly four thousand years old. It originated in about 1800 BC, with the Babylonians, who solved linear and quadratic equations much as we do today. They had no symbolic notation, so their equations had numerical coefficients, with their number system consisting of positive integers and rationals. Their solutions were prescriptive: do such and such and you will arrive at the answer. But the numerous repetitions of the same type of solution suggest that the procedure functioned as a standard algorithm.

The Greek Diophantus, the Indians, and the Chinese had various methods for solving linear and quadratic equations in the early centuries AD. The most influential Islamic work was likely al-Khwārizmī's *Kitab al-jabr wa'l-muqabalah* (*The Book of Restoring and Balancing*), of about 825 AD, in which he gave a classification and systematic treatment of quadratic equations. But he too had no symbolic notation, and, unlike the Indians and the Chinese, he considered as coefficients or roots only *positive* integers or rationals. Thus he deemed equations such as $3x+7=5$, $x^2=2$, and $x^2+1=0$ to be unsolvable.

Three major challenges deriving from the history just cited were:

i. to enlarge the number system to allow for the solution of equations such as those above,
ii. to extend the study of polynomial equations beyond those of the second degree, and
iii. to introduce a symbolic notation in order to develop a general theory of polynomial equations.

To a large extent these tasks were accomplished by the early seventeenth century. Mathematicians introduced symbolic notation in algebra; gradually began to use negative, real, and complex numbers, even if these were not rigorously defined; and succeeded in solving equations of degrees 3 and 4 "by radicals"—that is, they found a "formula" that expressed the roots of such equations in terms of rational operations (sums, differences, products, and quotients) on their coefficients and the extraction of roots, as in the quadratic formula (see ► Chapter 2). These were giant strides in the creation of what came to be known as "classical algebra". A crowning achievement of that subject was the Fundamental Theorem of Algebra (proved by Carl Friedrich Gauss in 1801), which in one of its versions says that every polynomial with complex coefficients has a complex root.

The Fundamental Theorem of Algebra asserts only the *existence* of roots, and so does not address the problem of *solving* polynomial equations—in particular, of solving them *by radicals*. This last challenge was met in the 1830s by Evariste Galois, an achievement which led to the introduction of groups (see ► Chapter 2 and [12]). Another prominent question, dealt with

» The scientist who first put forward in a systematic way the new propositions evoked by the problems sent to Pascal and Fermat, who gave the rules and who first made definitive the idea of mathematical expectation, was Christianus Huygens.

Huygens' book contained fourteen propositions, which were detailed solutions of problems dealing with games of chance. For example, two of the propositions were the Dice and Division Problems, presented by de Méré to Pascal. The ninth proposition discusses the Problem of Points involving an arbitrary number of players, and the twelfth proposition asks for the number of dice a player must use so that at least two sixes show up in a single throw.

The solutions of all the problems were carefully justified. The justifications were based on the fundamental notion of mathematical "expectation" (expected gain)—which Huygens was the first to define and highlight—rather than on the concept of probability, which is not mentioned. (One can define expectation in terms of probability *or* probability in terms of expectation [1, p. 165].) Huygens concluded his book with five challenging problems, which came to be named after him, and which enticed prominent mathematicians, including Jakob Bernoulli and Abraham De Moivre, to work on their solution. Here is the second problem [16, p. 25]:

» Three players, A, B, C, take twelve balls, eight of which are black and four white. They play on the following conditions: they are to draw blindfold, and the first who draws a white ball wins. A is to have the first turn, B the next, C the next; then A again, and so on. Determine the chances of the players.

Bernoulli solved the problem under several interpretations—for example, drawing the balls with or without replacement. See [9, p. 75] for Huygens' own solution of the division problem.

Huygens' book served as a text in probability—the only one available for the next fifty years, when it was incorporated, with commentary, as Part I of Jakob Bernoulli's *Ars Conjectandi* [2]. See the section below, as well as [3, 9] for details.

4.3 Jakob Bernoulli's *Ars Conjectandi (The Art of Conjecturing)*

The Bernoullis, a distinguished Swiss family, produced eight members who made significant contributions to mathematics. Most prominent among them were Jakob (1654–1705), his younger brother Johann (1667–1748), Johann's son Daniel (1700–1782), and Jakob's nephew Nikolaus (1687–1759). Jakob Bernoulli's important and influential book *Ars Conjectandi,* published posthumously in 1713, may be said to have completed the first period in the evolution of probability and given a thrust to the second [2]. The modern philosopher Ian Hacking gives details [8, p. 143]:

» Jacques Bernoulli's *Ars conjectandi* presents the most decisive conceptual innovations in the early history of probability. …[Upon its publication] probability came before the public with a brilliant portent of all the things we know about it now: its mathematical profundity, its unbounded practical applications…and its constant invitation for philosophizing. Probability had fully emerged.

The *Ars Conjectandi* comprises four parts. The first is a reprint of Huygens' *On Reckoning at Games of Chance,* with extensions and elaborations of his solutions. (Bernoulli was greatly

Jakob Bernoulli (1654–1705)

influenced by Huygens' book.) There are also solutions of the five problems which Huygens left as exercises. Part II contains a systematic account of "the doctrine of permutations and combinations", as Bernoulli called it, including what came to be known as the Bernoulli numbers [5]; and Part III applies the previous work to solve a series of games more challenging than those considered in Huygens' treatise. It was in Part IV, however, that Bernoulli made a fundamental advance in the subject by stating and proving a "Law of Large Numbers"—the first limit theorem of probability theory. He called it the "Golden Theorem" and considered the result a greater accomplishment than if he had shown how to square a circle. In the twentieth century, when stronger versions of Bernoulli's theorem had been proved, his result came to be known as the "Weak Law of Large Numbers".

Roughly, Bernoulli's Law of Large Numbers enables us to determine *experimentally* the probability of an event whose a priori probability is not known. For example, if there is an *unknown* number of black and white pebbles in an urn, the probability of drawing a white pebble from the urn can only be determined experimentally—by sampling. Thus, if in n identical trials an event occurs m times, and if n is very large, then m/n should be near the actual—a priori—probability of the event, and should get closer and closer to that probability as n gets larger and larger. See [9] for a precise mathematical statement of Bernoulli's Law of Large Numbers.

Bernoulli saw as the most important aspect of his book the application of his Law of Large Numbers to practical problems in civil, moral, and economic contexts. Eventually he ran out of time (it took him about twenty years to compose the *Ars Conjectandi*), and the task was left to his successors; but with his work probability began to make inroads into statistics, a process which greatly intensified over the next two centuries and more, resulting in an inseparable marriage of the two disciplines, which has become indispensable in many walks of life. We mentioned earlier that Bernoulli was the first to define and use the concept of probability. Here is his definition [2, p. 89]:

» *Probability* … is degree of certainty, and differs from the latter as a part differs from the whole…. One thing … is called *more probable* … than another if it has a larger part of certainty, even though in ordinary speech a thing is called probable only if its probability notably exceed one-half of certainty. I say *notably,* for what equals approximately half of certainty is called *doubtful* or undecided.

This is not very enlightening as a working definition, as it leaves unanswered the question of how to *compute* probabilities. Nevertheless, "Bernoulli's *Ars Conjectandi* … deserves to be

References

1. Artmann, B.: Euclid—the Creation of Mathematics. Springer, New York (1999)
2. Bell, E.T.: Men of Mathematics. Simon and Schuster, New York (1937)
3. Davis P.J., Hersh R., Marchisotto E.A.: The Mathematical Experience, Study Edition. Birkhäuser (1995). (orig. 1981)
4. Euclid: The Thirteen Books of Euclid's Elements, Translated, with extensive commentary, T. L. Heath. Dover, New York (1956)
5. Gans D.: An Introduction to NonEuclidean Geometry. Academic, New York (1973)
6. Gray J.: Ideas of Space: Euclidean, Non-Euclidean, and Relativistic, 2nd edn., Oxford University Press, New York (1989)
7. Gray J.: Worlds Out of Nothing: A Course in the History of Geometry in the 19th Century. Springer, New York (2007)
8. Greenberg M.J.: Euclidean and Non-Euclidean Geometries: Development and History, 4th edn. W. H. Freeman, New York (1981)
9. Kline M.: Mathematical Thought from Ancient to Modern Times. Oxford University Press, New York (1972)
10. Laugwitz D.: Bernhard Riemann, 1826–1866. Birkhäuser, New York (2008). (Translated by A. Shenitzer.)
11. Meschkowski H.: Noneuclidean Geometry. Academic, New York (1964). (Translated by A. Shenitzer.)
12. Niven I.: The roots of a quaternion. Am. Math. Mon. **49**, 386–388 (1942)
13. Trudeau R.J.: The Non-Euclidean Revolution. Birkhäuser, Boston (1987)
14. Wigner E.P.: The unreasonable effectiveness of mathematics in the natural sciences. Commun. Pure Appl. Math. **13**, 1–14 (1960)
15. Wolfe H.E.: Introduction to Non-Euclidean Geometry. Holt, Rinehart & Winston, Austen (1945)
16. Yandell B.H.: The Honors Class: Hilbert's Problems and Their Solvers. A K Peters, Natick (2002)

Further Reading

17. Browder F.E.: Does pure mathematics have a relation to the sciences? Am. Sci. **64**, 542–49 (1976)
18. Dyson F.J.: Mathematics in the physical sciences. Sci. Am. **211**, 129–46 (1964)
19. Eves H.: Great Moments in Mathematics (after 1650). Mathematical Association of America, Washington DC (1981)
20. Hadamard J.: The Psychology of Invention in the Mathematical Field. Dover, New York (1954)

Eugene Wigner (1902–1995)

We stand in awe and admire.

Problems and Projects

1. Prove that Euclid's fifth postulate P5 is equivalent—in the presence of his other four postulates—to Playfair's axiom, namely that through a point not on a given line there exists only one line parallel to the given line.
2. Outline Saccheri's approach to the proof of P5. See [5, 9, 15].
3. Show how Legendre "deduced" P5 from Euclid's other four postulates and indicate where Legendre went wrong. See [7, 15].
4. Examine Gauss' contribution to noneuclidean geometry, including the reasons why he did not publish in this area; describe his attitude towards Janos Bolyai. See [6–9].
5. Describe briefly what elliptic (noneuclidean) geometry is about. See [5–8, 10, 11, 15].
6. Describe the Klein model for noneuclidean (hyperbolic) geometry, and explain how it shows that hyperbolic geometry is consistent. See [8].
7. Discuss briefly Riemannian geometry and its relation to Einstein's theory of relativity. See [6–10].
8. Write a brief account of the life and some of the work of one of Bolyai, Lobachevski, Gauss, or Riemann. See [6, 7, 9, 10].
9. Explain what impact the creation of noneuclidean geometry had on the relationship of geometry to the physical world. See [6, 7, 9, 10, 15].
10. The invention of noneuclidean geometry had a substantial impact on the philosophy of Immanuel Kant. Explain. See [6–8, 10, 13, 15].
11. What was found deficient in Euclid's formulation of geometry and how was it set right? See [8, 15].
12. Find two results in mathematics which were introduced without any thought of being applied, yet which subsequently turned out to have important uses.

considered the founding document of mathematical probability", contends—not unjustly—Edith Dudley Sylla, who translated the book, with annotations, from Latin into English [2, p. vii].

4.4 De Moivre's *The Doctrine of Chances*

Another work on probability theory in the eighteenth century was De Moivre's *The Doctrine of Chances*—an outstanding text of the period. It began with a definition of probability [10, p. 646]:

» The Probability of an Event is greater, or less, according to the number of Chances by which it may either happen or fail.

De Moivre did very significant work in "pure" mathematics, but he was unable to obtain a university position, so he turned his efforts to the study of probability and life insurance. But he

» was rather old when he began his research in mathematics; and not until the age of 41 did he begin his work on probability theory. Nevertheless, he succeeded in becoming the leading probabilist from 1718 until his death, and he found one of his most important results, the normal approximation to the binomial distribution, in 1733, at the age of 66 [9, p. 401].

4.5 Laplace's *Théorie Analytique des Probabilités*

This is a seminal work on probability by one of the great mathematicians and scientists, Pierre Simon, Marquis de Laplace. It "summarized the results of the classical probability theory [that related to games of chance] and gave a decisive thrust to its further development" [15, p. 69].

Laplace believed that the theory of probability should be brought to bear on the social sciences, just as analysis has been brought to bear on the physical sciences. In support of this thesis he applied probability to decision theory, to the credibility of witnesses, and to insurance.

The 635-page *Théorie Analytique des Probabilités,* though deep and comprehensive, was a forbidding work, not easily accessible to the nonspecialist. So Laplace wrote a considerably shorter, "reader-friendly" book of 153 pages, with hardly any mathematical symbols, entitled *A Philosophical Essay on Probabilities* [12]. This deals essentially with the same subject matter as the *Théorie*. Its aim was to "present without the aid of analysis [that is, without the mathematical machinery] the principles and general results of this theory [probability], applying them to the most important questions of life, which are indeed for the most part only problems of probability" [12, p. 1]. The following is Laplace's definition of probability [12, p. 6]:

» The theory of chance ["chance" and "probability" were used interchangeably] consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose dominator is the number of all the cases possible.

Pierre Simon, marquis de Laplace
(1749–1827)

This definition was acceptable in the nineteenth century but is not satisfactory from a modern perspective; in particular, it does not accommodate infinite probability spaces.

4.6 Philosophy of Probability

We note that the several definitions of probability which we have given are rather wordy and seem less than satisfactory. In fact, as the historian Anders Hald asserts, "the concept of probability is an ambiguous one. It has gradually changed content, and at present it has many meanings [for example, objective and subjective probability; but see the next section], in particular in the philosophical literature" [9, p. 28].

So what is the nature of probability? Where did it come from? How do we describe/define it? These are largely philosophical questions, and they have been of concern mainly to philosophers. That is not surprising, since of course probability is closely connected to ideas such as causality and determinism. (Laplace too, in his *Philosophical Essay,* reflects on philosophical issues; see for example his Chapter IV, titled "Concerning Hope" [12].) Moreover, to make sense of the immense development of probability and its applications in the twentieth century, philosophers have introduced a number of "theories" of the subject, among them the classical theory, the logical theory, the subjective theory, the frequency theory, and the propensity theory [7].

4.7 Probability as an Axiomatic Theory

A number of outstanding mathematicians, among them Poisson, Gauss, Chebyshev, Markov, Bertrand, and Poincaré, made fundamental contributions to probability in the nineteenth century. Moreover, the subject had significant applications in the physical and social sciences. But it lacked foundations and was considered, according to Rényi, "a problematic discipline between mathematics and physics or philosophy" [15, p. 71]. Only in the early twentieth century did it begin to gain acceptance as a respectable branch of pure mathematics.

Bernhard Riemann (1826–1866)

(f) *Why is mathematics useful?*
For two thousand years, mathematics and the physical world were closely connected, the former serving as a model for aspects of the latter. This intimate relationship was fractured in the nineteenth century by the discovery of noneuclidean geometry. In particular, mathematical space and physical space became two distinct entities, with no evident connection between them.

Yet a tight linkage between mathematics and the physical world does exist. Mathematics abounds with examples of results and theories which were introduced without any thought of application yet which subsequently—a decade, a century, or a millennium later—turned out to be extremely useful. For example, matrices were introduced by Cayley in the 1850s simply as a useful algebraic notation, yet decades later they turned out to have numerous weighty uses. Another example: conic sections were introduced in ancient Greece to solve problems in pure mathematics but were used two thousand years later by Kepler in astronomy and by Galileo in mechanics. For a third example we cite Albert Einstein, who needed a Riemannian (noneuclidean) geometry, introduced in the 1850s, to formulate his theory of general relativity (in 1916).

How do we explain that "tight linkage" between mathematics and the physical world? The philosopher and mathematician Alfred North Whitehead found it paradoxical [9, p. 466]:

» The paradox is now fully established that the utmost abstractions are the true weapons with which to control our thought of concrete fact.

The Nobel Prize winner Eugene Wigner spoke famously of "the unreasonable effectiveness of mathematics in the natural sciences" [14, pp. 2, 7, 14]:

» The enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious and there is no rational explanation for it. …The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. …[It is] quite comparable in its striking nature to the miracle that the human mind can string a thousand arguments together without getting itself into contradictions or to the two miracles of the existence of laws of nature and of the human mind's capacity to divine them.

(c) *Axioms as self-evident truths*
There is little doubt that euclidean geometry was conceived by its creators as an idealization of physical space, its postulates suggested by, and abstracted from, physical experience. Euclid's postulates were thus deemed self-evident. With the creation of noneuclidean geometry, this view of the postulates could no longer be upheld. For if a postulate is self-evidently true, its negation would be false. What then are we to make of the contradictory fifth postulates of euclidean and hyperbolic geometry, each an integral "truth" in its respective geometry?

If euclidean and noneuclidean geometry are to coexist as mathematical systems, we must abandon the view of axioms as self-evident truths. Axioms are neither self-evident nor true. If Euclid's parallel postulate (say) was self-evident, how could its negation be a postulate in an equally consistent geometry? Moreover, if that postulate was true, then its negation would be false, and this would invalidate the consistency of noneuclidean geometry.

What then are axioms, if not self-evident truths? They are the starting *assumptions* of a mathematical theory—the "building blocks" of the theory. How to choose them so as to yield a useful theory is another matter.

(d) *What is geometry?*
It is important to realize that our *mathematical* conception of geometry must be divorced from its possible applicability to the physical world. And the notion that geometry represents truths about physical space must be abandoned. What then is geometry? It is a collection of various "geometries"—euclidean, hyperbolic, elliptic, projective, differential, algebraic, inversive, and so on. Each of these is a mathematical theory in its own right, based on its own set of assumptions (axioms) from which logical consequences (theorems) are deduced (proved). (An entirely different approach to geometry, via groups, was taken by Felix Klein in his Erlanger Program of 1872 [9].)

(e) *Relative truth*
A geometry is a set of logical consequences (theorems) of arbitrary (but consistent) assumptions (axioms) about meaningless entities (primitive terms), which we may designate as "points", "lines", and other "geometrical entities" (see ▶ Chapter 1). Since the primitive terms are not invested with meaning, neither are the axioms nor the theorems. So the theorems cannot be (absolutely) true. However, they are said to be *relatively true*—that is, true relative to the axioms of which they are consequences. Similar considerations apply to other mathematical structures defined by axioms. Here are two examples:

i. What is the sum of the angles of a triangle? The question as it stands is meaningless. It calls for another question: Is it a triangle in euclidean geometry?—in which case the sum of the angles of the triangle is 180°; or is it a noneuclidean triangle?—a hyperbolic triangle, whose angle sum is less than 180° or an elliptic triangle, with angle sum greater than 180°. So the answer depends on the *context*.
ii. Does the equation $x^2+1=0$ have any solutions, and if yes, how many? Again, the answer depends on the context: Over the reals the equation has no solutions, over the complex numbers it has two solutions, i and $-$i, and over the quaternions it has infinitely many—in fact, uncountably many—solutions; thus, $bi+\sqrt{(1-b^2)}\,j$ is a solution of $x^2+1=0$ for every real number b satisfying $-1\le b\le 1$ [12].

The issue of absolute versus relative truth is perhaps less important to the mathematician than to the philosopher. To the latter, mathematics provided, for two millennia, our sole example of absolute, indisputable, truth—until the 1830s, when that illusion was lost! See [3, 9].

Abstraction and axiomatization were hallmarks of mathematics in the first half of the twentieth century (see ▶ Chapter 1), so an axiomatic treatment of a branch of the subject seemed the proper approach to its study. David Hilbert was arguably the moving figure in inspiring and urging this view. In his celebrated 1900 address on Mathematical Problems, in which he singled out those that he thought should get the attention of research mathematicians of the twentieth century, the sixth problem asked for the axiomatization of probability (and of mechanics).

That task was accomplished in 1933 by the prominent Russian mathematician Andrey Nikolaevich Kolmogorov, with essential help from the recently created Lebesgue theory of measure and integration. For example, the expectation in this development of probability is a Lebesgue integral [15, p. 70].

The following are Kolmogorov's axioms for probability; they apply only to finite probability spaces [11, p. 2]:

Let **E** be a set whose elements we call "elementary events", and let **F** be a set of subsets of **E**, whose elements we call "random events". Let P be a function from **F** to $\mathbf{R}^*$, the set of all nonnegative real numbers. For each A ∈ **F**, we call P(A) the "probability" of the event A.

F is said to be a "field of probability" if the following axioms hold:

1. **F** is closed under the union, intersection, and difference of sets
2. **E** ∈ **F**
3. P(**E**) = 1
4. If A, B ∈ F such that $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

These then are the axiomatic foundations of finite probability theory. Kolmogorov extended this theory to infinite-dimensional probability spaces. In an 80-page booklet based on these axioms he discussed fundamental ideas of the subject, among them conditional probability, Bayes' theorem, random variables, mathematical expectation, independence, and the law of large numbers [11].

4.8 Conclusion

The creation of probability has been a "great moment" in mathematics [6]. Howard Eves (1911–2004) put it well [6, p. 9]:

> It is fascinating, and at the same time somewhat astonishing, to contemplate that mathematicians have been able to develop a science, namely the mathematical theory of probability, that establishes *rational laws* that can be applied to situations of *pure chance* [our emphases].

We shall give the last word to Laplace [12, pp. 195–196]:

> It is remarkable that a science which commenced with the consideration of games of chance should be elevated to the rank of the most important subjects of human knowledge. …there is no science more worthy of our meditations, and no more useful one could be incorporated in the system of public instruction.

Problems and Projects

1. Discuss the life and work of a mathematician encountered in this chapter who especially appealed to you.
2. Discuss Pascal's "Wager" concerning the existence of God. See [3, 4, 8, 9].
3. Discuss two or three paradoxes of probability theory.
4. Write a brief essay on John Graunt and his Mortality Tables. See [4, 9].
5. Write an essay on Pascal's *Treatise on the Arithmetical Triangle*. See [5, 6, 9].
6. Discuss some of Cardano's work in probability. See [8, 13, 14].
7. Give a numerical example of the Division Problem and explain how you would solve it. See [1, 4, 5, 9, 10].
8. Discuss aspects of the philosophy of probability. See [8, 9, 17].

References

1. Berlinghoff, W.P., Gouvêa, F.Q.: Math Through the Ages: A Gentle History for Teachers and others, expanded edition. Mathematical Association of America, Washington DC (2004)
2. Bernoulli, J.: The Art of Conjecturing (Ars Conjectandi), The Johns Hopkins University Press, Baltimore (2006). (Translated from the Latin (1713), with an introduction and notes, by E. D. Sylla.)
3. David, F.N.: Games, Gods, and Gambling. Hafner (1962)
4. Devlin, K.: The Unfinished Game: Pascal, Fermat, and the Seventeenth-Century Letter that Made the World Modern. Basic Books, New York (2008)
5. Edwards, A.W.F.: Pascal's Arithmetical Triangle. Oxford University Press, Oxford (1987)
6. Eves, H.: Great Moments in Mathematics (After 1650). Mathematical Association of America (1981)
7. Gillies, D.: Philosophical Theories of Probability. Routledge, London (2000)
8. Hacking, I.: The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference. Cambridge University Press, Cambridge (1975)
9. Hald, A.: History of Probability and Statistics and Their Applications Before 1750. Wiley, Hoboken (2003)
10. Katz, V.: A History of Mathematics: An Introduction, 3rd edn. Addison-Wesley, Boston (2009)
11. Kolmogorov, A.N.: Foundations of the Theory of Probability, 2nd edn. Chelsea, New York (1956)
12. Laplace, P.S.: A Philosophical Essay on Probabilities, Translated from the 6th French edn. Dover, Addison-Wesley, Boston (1951) (orig. 1819)
13. Ore, O.: Cardano, The Gambling Scholar. Dover, New York (1965)
14. Ore, O.: Pascal and the invention of probability theory. Amer. Math. Mon. **67,** 409–419 (1960)
15. Rényi, A.: Letters on Probability. Wayne State University Press, Detroit (1972)
16. Todhunter, I.: A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace. Chelsea (1949) (orig. 1865)
17. von Mises, R.: Probability, Statistics and Truth. Dover New York (1951)

Further Reading

18. Edwards, C.H.: The Historical Development of the Calculus. Springer, New York (1979)
19. Mahoney, M.: The Mathematical Career of Pierre de Fermat, 2nd edn. Princeton University Press, Princeton (1994)
20. Stigler, S.M.: Statistics on the Table: The History of Statistical Concepts and Methods. Harvard University Press, Harvard, Cambridge (1999)

geometry. Riemann defined a *new type* of noneuclidean geometry, called "elliptic geometry", in which there are *no* parallel lines and the sum of the angles of a triangle is *greater* than 180°. In fact, he introduced an infinite number of noneuclidean geometries, of arbitrary dimension, now known as "Riemannian". One of the many ideas in his remarkable paper was the use of differential methods in noneuclidean geometry. Somewhat later, Arthur Cayley and Felix Klein obtained euclidean *and* noneuclidean geometry (both hyperbolic and elliptic) as subgeometries of projective geometry. Thus did noneuclidean geometry acquire firmer foundations and enter the mainstream of mathematics. Geometry flowered in the nineteenth century! See [7].

7.5 Some Implications of the Creation of Noneuclidean Geometry

We consider a number of major issues arising from the discovery of noneuclidean geometry. At the Second International Congress of Mathematicians in Paris in 1900, in a talk on Mathematical Problems, David Hilbert referred to that breakthrough as one of the two "most suggestive and notable achievements of the [nineteenth] century" in the field he called "the principles of analysis and geometry" (the other being "the arithmetical formulation of the concept of the continuum") [16, p. 395].

(a) *Consistency*
We have described two noneuclidean geometries, hyperbolic and elliptic, which were developed on the basis of sets of axioms differing in some respects from those of euclidean geometry. But are we at liberty to propose an *arbitrary* set of axioms and proceed to create a discipline whose content is the set of logical consequences of those axioms? Yes, but *only* if the chosen axioms are *consistent*—that is, do not lead to a contradiction.

The creators of noneuclidean geometry felt confident about the consistency of their axioms, having derived a large body of theorems without arriving at a contradiction and having noted that past generations had failed to prove Euclid's P5. But convincing as such evidence was, it did not (of course) constitute a formal proof of consistency of the given geometry. To prove consistency, mathematicians devised the notion of a "model" of a geometry [8]. By constructing a euclidean model of noneuclidean geometry, they showed the *relative* consistency of noneuclidean geometry—namely, that this geometry is consistent *if* euclidean geometry is. Subsequently it was shown that euclidean geometry is consistent if noneuclidean geometry is. This established the relative consistency of one geometry with respect to the other. See [8].

(b) *Euclid is finally vindicated*
The consistency of hyperbolic geometry at last settled the two thousand-year-old question concerning a proof of Euclid's fifth postulate. It showed the *impossibility* of deducing the postulate from the remaining four postulates. For if that deduction *were* possible, P5 would be a theorem also in hyperbolic geometry, since the first four postulates of euclidean geometry are also postulates of hyperbolic geometry. But then the fifth postulates of hyperbolic and euclidean geometry would both be results in hyperbolic geometry, which would yield the inconsistency of that geometry. Moreover, the (relative) consistency of euclidean geometry showed that the *negation* of P5 cannot be proved from the other four. This established that Euclid's P5 is, as we now say, *independent* of his other four postulates [15].

Nikolai Ivanovich Lobachevsky (1792–1856)

Janos Bolyai (1802–1860)

Thus the two thousand-year search for a proof of Euclid's fifth postulate gave birth, around 1830, to a new geometry—noneuclidean. It was a momentous achievement. According to Donald Coxeter (1907–2003), the great twentieth-century geometer,

> The effect of the discovery of hyperbolic geometry on our ideas of truth and reality has been so profound that we can hardly imagine how shocking the possibility of a geometry different from Euclid's must have seemed in 1820. [8, p. xxv]

This intellectual revolution suffered the kind of pains attendant on all such upheavals. A generation passed before its acceptance by the mathematical community. Gauss' death in 1855 saw the publication of his diaries containing his thoughts on noneuclidean geometry, and his authority helped legitimize interest in the subject. Yet for the first forty years or so of its history, noneuclidean geometry still lacked wide acceptance. It had not been shown to be consistent, and it had not been related to other branches of mathematics or to physical phenomena.

The turning point came in 1868 with the publication of two papers, respectively by Eugenio Beltrami and Bernhard Riemann. Beltrami's paper established the *consistency* of hyperbolic

Calculus: From Tangents and Areas to Derivatives and Integrals

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics,* Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_5, © Springer Science+Business Media, LLC 2015

5.1 Introduction

The invention of calculus is one of the great intellectual and technical achievements of civilization. Calculus has served for three centuries as the principal quantitative tool for the investigation of scientific problems. It has given mathematical expression to such fundamental concepts as velocity, acceleration, and continuity, and to aspects of the infinitely large and infinitely small—notions that have formed the basis for much mathematical and philosophical speculation since ancient times. Modern physics and technology would be impossible without calculus. The most important equations of mechanics, astronomy, and the physical sciences in general are differential and integral equations—outgrowths of the calculus of the seventeenth century. Other major branches of mathematics derived from calculus are real analysis, complex analysis, and calculus of variations. Calculus is also fundamental in probability, topology, Lie group theory, and aspects of algebra, geometry, and number theory. In fact, mathematics as we know it today would be inconceivable without the ideas of calculus.

Isaac Newton and Gottfried Wilhelm Leibniz independently invented calculus during the last third of the seventeenth century. But their work was neither the beginning of the story nor its end. Practically all of the prominent mathematicians of Europe around 1650 could solve many of the problems in which elementary calculus is now used—but providing their procedures with rigorous foundations required two more centuries.

The infinitely small and the infinitely large—in one form or another—are essential in calculus. In fact, they are among the features which most distinguish that branch of mathematics from others. They have appeared throughout the history of calculus in various guises: infinitesimals, indivisibles, differentials, "evanescent" quantities, moments, infinitely large and infinitely small magnitudes, infinite sums, and power series. Also they have been fundamental at both the technical and conceptual levels—as underlying tools of the subject and as its foundational underpinnings. We will give examples of these manifestations of the infinite in the earlier evolution of calculus (seventeenth and eighteenth centuries).

5.2 Seventeenth-Century Predecessors of Newton and Leibniz

The Renaissance (ca. 1400–1600) saw a flowering and vigorous development of the visual arts, literature, music, the sciences, and—not least—mathematics. It witnessed the decisive triumph of positional decimal arithmetic, the introduction of algebraic symbolism, the solution by radicals of the cubic and quartic, the free use if not full understanding of irrational numbers, the introduction of complex numbers, the rebirth of trigonometry, the establishment of a relationship between mathematics and the arts through perspective drawing, and a revolution in astronomy, later to prove of great significance for mathematics. A number of these developments were necessary prerequisites for the rise of calculus, as was the invention of analytic geometry by René Descartes and by Pierre de Fermat in the early decades of the seventeenth century (see ► Chapter 3).

The Renaissance also saw the full recovery and serious study of the mathematical works of the Greeks, especially Archimedes' masterpieces. His calculations of areas, volumes, and centers of gravity were an inspiration to many mathematicians of that period. Some went beyond Archimedes in attempting systematic calculations of the centers of gravity of solids. But they used the classical "method of exhaustion" of the Greeks, which was conducive neither to the discovery of results nor to the development of algorithms. The temper of the times was such that most mathematicians were far more interested in results than in proofs; rigor, declared Bonaventura Cavalieri in the 1630s, "is the concern of philosophy and not of geometry [mathematics]" [10, p. 383]. To obtain results, mathematicians devised new methods for the solution of calculus-type problems. These were based on geometric, algebraic, and arithmetic ideas, often in interplay. We give two examples.

Cavalieri

A major tool for the investigation of calculus problems was the notion of an *indivisible*. This idea—in the form, for example, of an area as composed of a sum of infinitely many parallel lines, regarded as atomistic—was embodied in Greek physical theory and was also part of medieval scientific thought. Mathematicians of the seventeenth century fashioned indivisibles into a powerful tool for the investigation of area and volume problems.

Indivisibles were used in calculus by Galileo and others in the early seventeenth century, but it was Cavalieri who, in his influential *Geometry of Indivisibles* of 1635, shaped a vague concept into a useful technique for the determination of areas and volumes. His strategy was to consider a geometric figure to be composed of an infinite number of indivisibles *of lower dimension*. Thus a surface consists of an infinite number of equally spaced parallel lines, and a solid of an infinite number of equally spaced parallel planes. The procedure for finding the area (or volume) of a figure is to compare it to a second figure of equal height (or width), whose area (or volume) is known, by setting up a one-to-one correspondence between the indivisible elements of the two figures and using "Cavalieri's Principle": if the corresponding indivisible elements are always in a given ratio, then the areas (or volumes) of the two figures are in the *same* ratio. For example, it is easy to show that the ordinates of the ellipse $x^2/a^2+y^2/b^2=1$ are to the corresponding ordinates of the circle $x^2+y^2=a^2$ in the ratio b:a (see ◘ Figure 5.1), hence the area of the ellipse $=(b/a)\times$ the area of the circle $=\pi ab$.

Fermat

Fermat was the first to tackle systematically the problem of tangents. In the 1630s he devised a method for finding tangents to any polynomial curve. The following example illustrates his approach.

Suppose we wish to find the tangent to the parabola $y=x^2$ at some point (x, x^2) on it. Let $x+e$ be a point on the x-axis and let s denote the "subtangent" to the curve at the point (x, x^2) (see ◘ Figure 5.2). Similarity of triangles yields $x^2/s=k/(s+e)$. Fermat notes that k is "adequal" to $(x+e)^2$, presumably meaning "as nearly equal as possible", although he does not say so. Writing this as $k\cong(x+e)^2$, we get $x^2/s\cong(x+e)^2/(s+e)$. Solving for s we have $s\cong ex^2/[(x+e)^2$

of a new—"noneuclidean"—geometry. Saccheri was on the verge of its discovery, but he would not—could not—accept his own results, because they contradicted propositions in euclidean geometry. This attitude was a barrier whose overcoming would require *not* a mathematical but a *psychological* breakthrough; and this Saccheri could not achieve. Perhaps, as Wolfgang Bolyai (the father of one of the inventors of noneuclidean geometry) claimed, "mathematical discoveries, like springtime violets in the woods, have their season which no human can hasten or retard" [2, p. 263].

7.4 The Discovery (Invention) of Noneuclidean Geometry

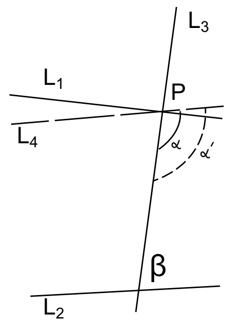
In 1763 a German student, G. S. Klügel, submitted a Ph.D. dissertation that found flaws in 28 different supposed proofs of the parallel postulate, and in 1766 Johann Lambert made further interesting discoveries along the lines of Saccheri. But the problem of the parallel postulate, still unresolved, was not at the centre of attention of eighteenth-century mathematics; the major concerns at the time were in analysis. It is only towards the beginning of the nineteenth century that we witness a revival of interest in geometry. In this context, Ferdinand Schweikart developed "astral geometry" in 1807 and Franz Taurinus a "logarithmic-spherical geometry" in 1826—both notable forerunners of noneuclidean geometry. Nevertheless, the following lament from Wolfgang Bolyai in the 1820s suggests that the dilemma posed by the parallel postulate still seemed far from resolved [11, p. 31]:

> » It is unbelievable that this stubborn darkness, this eternal eclipse, this flaw in geometry, this eternal cloud on virgin truth can be endured.

The German Carl Friedrich Gauss, the Hungarian Janos Bolyai, and the Russian Nikolai Lobachevsky are considered the independent inventors of noneuclidean geometry, although Gauss did not publish his researches in this field. These three mathematicians were the first to develop—*consciously and systematically*—a new geometry, which they regarded as logically consistent, and whose theorems included many of the "strange" results arrived at in past generations. Its point of departure was the acceptance of Euclid's first four postulates but the *replacement* of the fifth by an *opposed* "parallel" postulate, namely that through a point not on a given line there is *more* than one line parallel to the given line. The body of theorems derived as logical consequences of these postulates came to be known as "noneuclidean geometry" (later as "hyperbolic geometry"). Here are some of those theorems:

1. The sum of the angles of a triangle is less than 180°. It follows, in particular, that rectangles do not exist in this geometry.
2. The sum of the angles of a triangle varies with the area of the triangle—the larger the area, the smaller the angle sum.
3. Similar triangles are necessarily congruent.
4. Two distinct lines cannot be equidistant.
5. A line may intersect one of two parallel lines without intersecting the other.
6. The ratio of the circumference to the diameter of a circle is larger than π. Moreover, the ratio increases as the area of the circle increases.

Figure 7.2 Proclus' proof of the 5th postulate



logically equivalent—that is, in the presence of the other four postulates, this statement and P5 imply each other. And of course one must not use an equivalent of P5 to *prove* P5!

Similar attempts to prove P5 were made over the next fourteen centuries. All failed for essentially the same reason as Proclus', namely the dependence of the proof on an assumption—implicit or explicit—which seemed obvious but which was subsequently shown to be equivalent to P5. Below are some examples of such equivalents (with the names and dates of their originators, where known).

a. Two parallel lines are equidistant (Posidonius, first century BC).
b. If a line intersects one of two parallel lines, it also intersects the other (Proclus, fifth century AD).
c. There exists a rectangle (Nasir-Eddin, thirteenth century).
d. Given a triangle, we can construct a similar triangle of any size (John Wallis, seventeenth century).
e. Through a point not on a given line, there exists only one parallel to the given line (John Playfair, eighteenth century).
f. Three noncollinear points always lie on a circle (Legendre, end of eighteenth century).
g. Two lines parallel to a third line are parallel to each other.
h. The locus of all points equidistant from a straight line is a straight line.

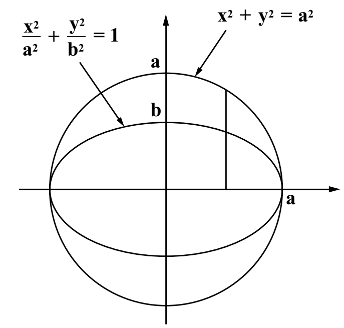
We should point out that because of statement (e), Euclid's fifth postulate is often referred to as the "parallel postulate".

Girolamo Saccheri, a Jesuit priest, made an important departure from the line of reasoning of his predecessors in his attempt to prove P5. In 1733 he wrote a book entitled *Euclid Vindicated of Every Flaw*, in which he assumed the *negation* of P5 and tried to arrive at a contradiction. Saccheri deduced a number of "strange" results, among them that the sum of the angles of a triangle is less than 180°, and that a line can approach another line asymptotically. He claimed that the latter result was "repugnant to the nature of the straight line" [8, p. 218]. He therefore concluded—erroneously—that Euclid's P5 must hold, namely, that it is a *theorem* derivable from his other four postulates.

Such results might indeed have appeared strange to anyone exposed exclusively to euclidean geometry, but they were not logically contradictory. In fact, they were to form the elements

Bonaventura Cavalieri (1598–1647)

Figure 5.1 Area of an ellipse

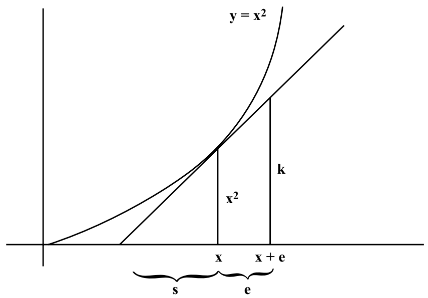


$-x^2] = ex^2/e(2x+e) = x^2/(2x+e)$. It follows that $x^2/s \cong 2x+e$. Note that x^2/s is the slope of the tangent to the parabola at (x, x^2). Fermat now "deletes" the e and claims that the slope of the tangent is 2x.

Fermat's method was severely criticized by some of his contemporaries, notably Descartes. They objected to his introduction and subsequent suppression of the "mysterious e". Dividing by e meant regarding it as not zero—but discarding e implied that it *was* zero. This is inadmissible, they rightly claimed. But Fermat's mysterious e embodied a crucial idea: the giving of a "small" increment to a variable. And it cried out for the limit concept, which was formally introduced only about two hundred years later. Fermat, however, considered his method to be purely *algebraic*.

The above examples give us a glimpse of the near-century of vigorous investigations in calculus prior to the work of Newton and Leibniz. Mathematicians plunged boldly into almost virgin territory—the mathematical infinite—where a more critical age might have feared to tread. They produced a multitude of powerful, if nonrigorous, infinitesimal techniques for the solution of area, volume, and tangent problems. What, then, was left for Leibniz and Newton to do?

Figure 5.2 Finding the tangent to a parabola



5.3 Newton and Leibniz: The Inventors of Calculus

In the first two thirds of the seventeenth century mathematicians solved calculus-type problems, but they lacked a general framework in which to place them. This was provided by Newton and Leibniz. Specifically, they

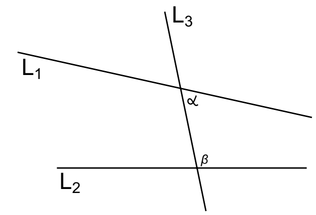
a. invented the general concepts of derivative and integral—though not in the form we see them today. For example, it is one thing to compute areas of curvilinear figures and volumes of solids using ad hoc methods, but quite another to recognize that such problems can be subsumed under a single concept, namely the integral.
b. recognized differentiation and integration as inverse operations. Although several mathematicians before Newton and Leibniz noted the relation between tangent and area problems, mainly in specific cases, the clear and explicit recognition, in its complete generality, of what we now call the Fundamental Theorem of Calculus belongs to Newton and Leibniz.
c. devised a notation and developed algorithms to make calculus a powerful computational instrument.
d. extended the range of applicability of the methods of calculus. While in the past those methods were applied mainly to polynomials, often only of low degree, they were now applicable to "all" functions, algebraic and transcendental.

And now to some examples of the calculus as developed by Newton and by Leibniz. We should note that theirs is a calculus of *variables*—which Newton calls "fluents"—and equations relating these variables; it is *not* a calculus of *functions*. The notion of function as an explicit mathematical concept arose only in the early eighteenth century.

Newton

Newton considered a curve to be "the locus of the intersection of two moving lines, one vertical and the other horizontal. The x and y coordinates of the moving points are then functions of the time t, specifying the locations of the vertical and horizontal lines respectively" [4, p. 193]. Newton's basic concept is that of a "fluxion", denoted by $\dot{x}$; it is the instantaneous rate of change (instantaneous velocity) of the fluent x—in our notation, dx/dt. The instantaneous velocity is not defined, but is taken as intuitively understood. Newton aims rather to show how to *compute* $\dot{x}$.

Figure 7.1 Euclid's 5th postulate (P5)



be "an epoch-making statement" [15, p. 17]—was apparently singled out for special attention from earliest times: it took considerably longer to state than the other four, and was not nearly so self-evident. Euclid himself may have felt uneasy about this postulate, for his first use of it comes after he deduced twenty-eight propositions from only the other four [1, 4].

This fifth postulate would play a large role in subsequent history, and so it appears often in the story that we tell below; for variety and brevity we shall sometimes refer to it as "P5".

Proclus, a Greek philosopher and mathematician whose works are among our main sources of information on Greek geometry, stated the dilemma thus in his *Commentary on Euclid's Elements* [8, p. 210]:

> This [P5] ought even to be struck out of the Postulates altogether; for it is a statement involving many difficulties. … The statement that since the two lines converge more and more as they are produced will eventually meet is plausible but not necessary.

To substantiate the last statement, Proclus gave the example of a hyperbola and its asymptotes, and he consequently proposed the following [8, p 210]:

> It is then clear from this that we must seek a proof of the present theorem, and that it is alien to the special character of postulates.

7.3 Attempts to Prove the Fifth Postulate

Proclus himself offered such a proof:

Let L_3 be a line intersecting the lines L_1 and L_2 such that $\alpha + \beta < 180°$; we want to prove that L_1 and L_2 intersect (see Figure 7.2). Since $\alpha + \beta < 180°$, draw a line L_4 through P (the point of intersection of L_2 and L_3) such that $\alpha' + \beta = 180°$. It follows that L_4 and L_2 are parallel. (This is Proposition 28 of Euclid's *Elements*; it is proved without the use of the fifth postulate P5.) Now Proclus argued that since L_1 intersects L_4 (namely at P), it must also intersect L_2, basing himself on the allegedly obvious fact that if a line intersects one of two parallel lines it must intersect the other. Thus L_1 and L_2 intersect, which completes the proof of P5.

The problem with this proof is that while the statement "if a line intersects one of two parallel lines it must intersect the other" may be more self-evident than P5, the two are in fact

Noneuclidean Geometry: From One Geometry to Many

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics*, Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_7, © Springer Science+Business Media, LLC 2015

7.1 Introduction

For over two millennia there was only one geometry—the one we now call "euclidean". Since it was believed to consist of truths abstracted from the physical world, it was taken to be the only *possible* geometry. But there was a perceived blemish: one of this geometry's postulates was not as self-evident as the others. So mathematicians tried to derive it as a *proposition* (theorem) from the remaining postulates. Such attempts were made over many centuries, by, among others, John Wallis and Adrien-Marie Legendre—but to no avail. Early in the nineteenth century two young, little-known mathematicians, the Hungarian Janos Bolyai and the Russian Nikolai Lobachevsky, dared the establishment by proposing (independently) a new geometry, different from Euclid's yet mathematically just as valid—a "noneuclidean" geometry, based on the replacement of Euclid's troublesome postulate by an alternative. The discovery of that geometry had a deep impact on mathematics, on science, and on philosophy.

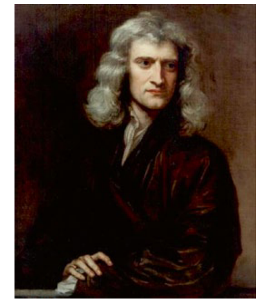
7.2 Euclidean Geometry

Our story begins with Euclid, ca. 300 BC. His singular achievement was to bring under the roof of the axiomatic method (see ▸ Chapter 1) the fundamentals of geometry as they had developed over the previous three centuries. The resulting grand opus, titled *Elements*, became an ideal of exposition in mathematics, and in other subjects, for more than two millennia. It contains over 400 propositions, logically deduced from only five postulates. The postulates are:

1. A straight line may be drawn between any two points.
2. A straight line segment may be produced indefinitely.
3. A circle may be drawn with any given point as centre and any given radius.
4. All right angles are equal.
5. If a straight line intersects two other straight lines lying in a plane, and if the sum of the interior angles thus obtained on one side of the intersecting line is less than two right angles, then the straight lines will eventually meet, if extended sufficiently, on the side on which the sum of the angles is less than two right angles.

In terms of the diagram below (Figure 7.1), postulate 5 states that if L_3 cuts L_1 and L_2 such that $\alpha + \beta$ 180°, then L_1 and L_2, if produced, will intersect to the right of L_3.

These postulates were viewed by the Greeks as "self-evident truths", since they were deemed to be idealizations of physical space, and therefore they were assumed to require no proof. But Euclid's fifth postulate—which the mathematician Cassius Keyser (1862–1947) judged to



Isaac Newton (1642–1727)

The following is an example of Newton's computation of the tangent to a curve with equation $x^3 - ax^2 + axy - y^3 = 0$ at an arbitrary point (x, y) on the curve. He lets o be an infinitesimal period of time. Then $\dot{x}o$ and $\dot{y}o$ are infinitesimal increments in x and y, respectively. (For, we have distance = velocity × time = $\dot{x}o$ or $\dot{y}o$, assuming with Newton that the instantaneous velocities $\dot{x}$ and $\dot{y}$ of the point (x, y) moving along the curve remain constant throughout the infinitely small time interval o.) Newton calls $\dot{x}o$ and $\dot{y}o$ *moments*, a "moment" of a fluent being the amount by which it increases in an infinitesimal time period. An infinitesimal was not formally defined, but was understood to be an "infinitely small" quantity, less than any finite quantity but not zero. Thus, $(x + \dot{x}o,\ y + \dot{y}o)$ is a point on the curve infinitesimally close to (x, y). In Newton's words: "Soe y^t if y^e described lines [coordinates] bee x and y, in one moment, they will bee $x + \dot{x}o$ and $y + \dot{y}o$ in y^e next" [4, p. 193]. Substituting $(x + \dot{x}o,\ y + \dot{y}o)$ into the original equation and simplifying by deleting $x^3 - ax^2 + axy - y^3$ (which equals zero) and dividing by o, we get:

$$3x^2\dot{x} - 2ax\dot{x} + ay\dot{x} + ax\dot{y} - 3y^2\dot{y} + 3x\dot{x}^2o - a\dot{x}^2o + a\dot{x}\dot{y}o - 3\dot{y}y^2o + \dot{x}^3o^2 - \dot{y}^3o^2 = 0.$$

Newton now discards the terms involving o, noting that they are "infinitely lesse" than the remaining terms. This yields an equation relating x and y, namely

$$3x^2\dot{x} - 2ax\dot{x} + ay\dot{x} + ax\dot{y} - 3y^2\dot{y} = 0.$$

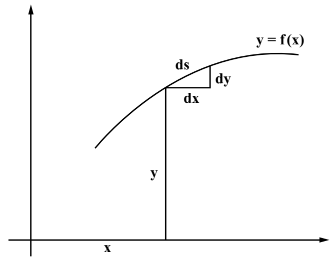
From this relationship we can get the slope of the tangent to the given curve at any point (x, y):

$$\dot{y} / \dot{x} = \frac{3x^2 - 2ax + ay}{3y^2 - ax}.$$

This procedure is quite general, Newton notes, and it enables him to obtain the slope of the tangent to *any* algebraic curve.

The problem of what to make of the "o's"—the "ghosts of departed quantities" [4, p. 294]—remained, according to Bishop George Berkeley, who launched a famous critique. Are they zero? Finite quantities? Infinitely small? Newton's dilemma was not unlike Fermat's a half-century earlier.

Figure 5.3 Leibniz' characteristic triangle



Leibniz

Leibniz' ideas on calculus evolved gradually, and like Newton, he wrote several versions, giving expression to his ripening thoughts. Central to all of them is the concept of "differential", although that notion had different meanings for him at different times.

Leibniz viewed a "curve" as a polygon with infinitely many sides, each of infinitesimal length. (Recall that the Greeks conceived a circle in just that way.) With such a curve is associated an infinite (discrete) sequence of abscissas x_1, x_2, x_3, …, and an infinite sequence of ordinates y_1, y_2, y_3, …, where (x_i, y_i) are the coordinates of the points of the curve.

The difference between two successive values of x is called the "differential" of x and is denoted by dx; similarly for dy. The differential dx is a fixed nonzero quantity, infinitely small in comparison with x—in effect, an infinitesimal. There is a *sequence* of differentials associated with the curve, namely the sequence of differences $x_i - x_{i-1}$ associated with the abscissas x_1, x_2, x_3, … of the curve [4, pp. 258, 261].

The sides of the polygon constituting the curve are denoted by ds—again, there are infinitely many such infinitesimal ds's. This gives rise to Leibniz' famous "characteristic triangle" with infinitesimal sides dx, dy, ds satisfying the relation $(ds)^2 = (dx)^2 + (dy)^2$ (see Figure 5.3). The side ds of the curve (polygon) is taken as coincident with the tangent to the curve (at the point x). Leibniz put it thus [9, pp. 234–235]:

> We have only to keep in mind that to find a *tangent* means to draw a line that connects two points of the curve at an infinitely small distance, or the continued side of a polygon with an infinite number of angles, which for us takes the place of the *curve*. This infinitely small distance can always be expressed by a known differential like d*s*.

The slope of the tangent to the curve at the point (x, y) is thus dy/dx—an actual *quotient* of differentials, which Leibniz calls the "differential quotient" (Figure 5.3).

Here are two further examples of his calculus. To discover and "prove" the product rule for differentials, he proceeds as follows:

$d(xy) = (x + dx)(y + dy) - xy = xy + xdy + ydx + (dx)(dy) - xy = xdy + ydx.$ He omits (dx)(dy), noting that it is "infinitely small in comparison with the rest" [4, p. 255].

As a second example, Leibniz finds the tangent at a point (x, y) to the conic $x^2 + 2xy = 5$: Replacing x and y by x + dx and y + dy, respectively, and noting that (x + dx, y + dy) is a point on the conic "infinitely close" to (x, y), we get

$$(x + dx)^2 + 2(x + dx)(y + dy) = 5 = x^2 + 2xy.$$

3. Write brief essays on the lives and work of two of Fermat, Euler, Gauss, Dedekind.
4. Show that $Z_{-5} = \{a + b\sqrt{-5} : a, b \in Z\}$ is not a UFD. See [1, 16].
5. Determine all gaussian primes. See [1, 3, 17].
6. Write a brief essay on Diophantus, addressing both his algebraic and number-theoretic work, and discussing his influence. See [2, 6, 10].
7. Discuss Lagrange's solution of the Pell equation, $x^2 - dy^2 = 1$, d a positive integer, noting his use of "foreign objects" in number theory. See [1, 3, 5, 15].
8. Write an essay on Bachet, Frenicle, and Mersenne, the scientists who were Fermat's correspondents.
9. Write an essay on the factorization of ideals in rings of integers of quadratic fields. See [1, 4, 5, 11, 13].
10. Discuss the law of quadratic reciprocity, and the law of biquadratic reciprocity. See [1, 7, 17].

References

1. Adams, W.W., Goldstein, L.J.: Introductions to Number Theory. Prentice-Hall, Englewood Cliffs (1976)
2. Bashmakova, I.G.: Diophantus and Diophantine Equations (translated from the Russian by A. Shenitzer). Mathematical Association of America, Washington DC (1997)
3. Bolker, E.D.: Elementary Number Theory: An Algebraic Approach. W. A. Benjamin, New York (1970)
4. Edwards, H.M.: Fermat's Last Theorem: A Genetic Introduction to Algebraic Number Theory. Springer, New York (1977)
5. Goldman, J.R.: The Queen of Mathematics: A Historically Motivated Guide to Number Theory. A K Peters, Wellesley (1998)
6. Heath, T.L.: Diophantus of Alexandria: A Study in the History of Greek Algebra, 2nd ed. Dover, New York (1964) (Contains a translation into English of Diophantus' *Arithmetica*, a 130-page Introduction to Diophantus' and related work, and a 60-page Supplement on Fermat's number-theoretic work.)
7. Ireland, K., Rosen, M.: A Classical Introduction to Modern Number Theory. Springer, New York (1972)
8. Katz, V.J.: A History of Mathematics, 3rd ed. Addison-Wesley, Hoboken (2009)
9. Kleiner, I.: A History of Abstract Algebra. Birkhäuser, Basel (2007)
10. Ore, O.: Number Theory and its History. McGraw-Hill, New York (1948)
11. Pollard, H., Diamond, H.G.: The Theory of Algebraic Numbers, 2nd ed. Mathematical Association of America, Washington DC (1975)
12. Ribenboim, P.: 13 Lectures on Fermat's Last Theorem. Springer, New York (1979)
13. Robinson, A.: Numbers and Ideals. Holden-Day, San Francisco (1965)
14. Singh, S.: Fermat's Enigma: The Quest to Solve the World's Greatest Mathematical Problem. Penguin, New York (1997)
15. Stark, H.M.: An Introduction to Number Theory. MIT Press, Cambridge (1978)
16. Stillwell, J.: Elements of Number Theory. Springer (2000)
17. Stillwell, J.: Mathematics and Its History. Springer, New York (2010) (orig. 1989)
18. Weil, A.: Number Theory: An Approach Through History, from Hammurapi to Legendre. Birkhäuser, Boston (1984)

Further Reading

19. Dickson, L.E.: History of the Theory of Numbers, Vol. 3. Chelsea, New York (1966) (orig. 1919, 1920, and 1923)
20. Kleiner, I.: From Fermat to Wiles: Fermat's last theorem becomes a theorem. Elem. Math. 55 , 19–37 (2000)

Richard Dedekind (1831–1916)

Q(a)". (For example, $\sqrt{15}+3$ is an integer of $Q(\sqrt{15})$, since it is a root of the monic polynomial x^2-6x-6.) In fact, Dedekind *defined* a ring—the first such definition. The I(a) are integral domains, but they are not in general ufds.

The I(a) are vast generalizations of the domains of integers that were considered in this chapter: the gaussian integers, the cyclotomic integers, the quadratic integers, and of course the ordinary integers. They are also, Dedekind determined, the appropriate domains in which to formulate and prove a UFT. This turned out to be the following: *Every nonzero and noninvertible ideal in* I(a) *is a unique product of prime ideals.*

These I(a) are examples of "dedekind domains", which play an essential role in (algebraic) number theory (cf. unique factorization domains, which play an essential role in "elementary" number theory).

6.6 Conclusion

To summarize the events that we have been describing: after more than two thousand years in which number theory meant the study of properties of the (positive) integers, its scope became enormously enlarged. One could no longer use the term "integer" with impunity: it had to be qualified—a "rational" (ordinary) integer, a gaussian integer, a cyclotomic integer, a quadratic integer, or any one of an infinite species of other (algebraic) integers—the various I(a). Moreover, powerful new algebraic tools were introduced and brought to bear on the study of these integers—fields, commutative rings, unique factorization domains, ideals, prime ideals, and Dedekind domains. A new subject—algebraic number theory—had emerged, vitally important to this day.

Problems and Projects

1. Supply the steps needed to make Euler's solution of the Bachet equation $x^2+2=y^3$ rigorous. See [1, 3, 15, 16].
2. Solve $x^2+y^2=z^2$ in integers (that is, find all Pythagorean triples) using the ideas of this chapter, that is, factoring the left side of the equation and proceeding as in Euler's solution of the Bachet equation. See [1, 17].

Gottfried Wilhelm Leibniz (1646–1716)

Simplifying, and discarding (dx)(dy) and $(dx)^2$, which are assumed to be negligible in comparison with dx and dy, yields $2xdx+2xdy+2ydx=0$. Dividing by dx and solving for dy/dx gives $dy/dx=(-x-y)/x$. This is of course what *we* would get by writing $x^2+2xy=5$ as $y=(5-x^2)/2x$ and differentiating this *functional* relation. (Recall that Leibniz' calculus predates the emergence of the function concept.)

We see in these examples how Leibniz' choice of a felicitous notation enabled him to arrive very quickly at reasonable convictions, if not rigorous proofs, of important results. His symbolic notation served not only to *prove* results but also greatly facilitated their *discovery*.

5.4 The Eighteenth Century: Euler

Brilliant as the accomplishments of Newton and Leibniz were, their respective versions of calculus consisted largely of loosely connected methods and problems, and were not easily accessible to the mathematical public, such as that was. The first systematic introduction to the Leibnizian differential calculus was given in 1696 by Guillaume de L'Hospital in his text *The Analysis of the Infinitely Small, for the Understanding of Curved Lines*. Calculus was further developed during the early decades of the eighteenth century, especially by the Bernoulli brothers Jakob and Johann. Several books appeared during this period, but the subject lacked focus. The main contemporary concern of calculus was with the geometry of curves—tangents, areas, volumes, and lengths of arcs (cf. the title of L'Hospital's text). Of course Newton and Leibniz introduced an algebraic apparatus, but its motivation and the problems to which it was applied were geometric or physical, having to do with curves. In particular, this was (as we already noted) a calculus of variables related by equations rather than a calculus of functions.

A fundamental conceptual breakthrough was achieved by Euler around the mid-eighteenth century. This was to make the concept of *function* the centerpiece of calculus. Thus calculus is not about curves, asserted Euler, but about functions. The derivative and the integral are not merely abstractions of the notions of tangent or instantaneous velocity on the one hand and of area or volume on the other—they are the basic concepts of calculus, to be investigated in their own right. But mathematicians of the eighteenth century did not readily embrace this centrality of functions, especially since variables seemed to serve them well.

Leonhard Euler (1707–1783)

Power series played a fundamental role in the calculus of the seventeenth and eighteenth centuries, especially in Newton's and Euler's. They were viewed as infinite polynomials with little, if any, concern for convergence. The following is an example of Euler's derivation of the power-series expansion of sin*x*, employing infinitesimal tools with great artistry [4, p. 235]:

Use the binomial theorem to expand the left-hand side of the identity $(\cos z + i\sin z)^n = \cos(nz) + i\sin(nz)$, and equate the imaginary part to $\sin(nz)$. We then get:

$$\sin(nz) = n(\cos z)^{n-1}(\sin z) - \left[n(n-1)(n-2)/3!\right](\cos z)^{n-3}(\sin z)^3 \tag{5.1}$$
$$+\left[n(n-1)(n-2)(n-3)(n-4)/5!\right](\cos z)^{n-5}(\sin z)^5 - \ldots.$$

Now let n be an infinitely large integer and z an infinitely small number (Euler sees no need to explain what these are). Then

$$\cos z = 1, \sin z = z, n(n-1)(n-2) = n^3, n(n-1)(n-2)(n-3)(n-4) = n^5 \ldots.$$

(again no explanation from Euler, although of course we can surmise what he had in mind). Equation 5.1 now becomes

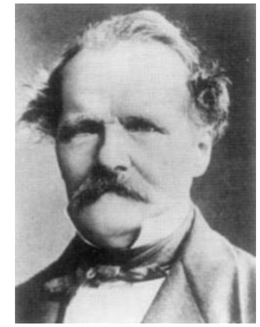
$$\sin(nz) = nz - \left(n^3z^3\right)/3! + \left(n^5z^5\right)/5! - \ldots.$$

Let now $nz = x$. Euler claims that x is finite since n is infinitely large and z infinitely small. This finally yields the power-series expansion of the sine function:

$$\sin x = x - x^3/3! + x^5/5! - \ldots$$. It takes one's breath away!

This formal, algebraic style of analysis, used so brilliantly by Euler and practiced by most eighteenth-century mathematicians, is astonishing. It accepted as articles of faith that what is true for convergent series is true for divergent series, what is true for finite quantities is true for infinitely large and infinitely small quantities, and what is true for polynomials is true for

Ernst Eduard Kummer (1810–1893)



Kummer's result was quite a feat, considering that during the previous two centuries FLT had been proved for only three primes. Further crucial progress would require another century and more.

Kummer's brilliant work went much beyond its application to FLT. Its main focus was the study of reciprocity laws (see earlier comments in this section). One of its major achievements was to "rescue" unique factorization (see above) in the domains $C_p = \{a_0 + a_1w + a_2w^2 + \ldots + a_{p-1}w^{p-1} : a_i \in Z\}$ of cyclotomic integers. He did this by showing that every nonzero, noninvertible element of C_p is a unique product of "ideal primes".

Kummer's work left important questions unanswered:

(i) What *is* an "ideal prime" anyway? This central concept in his work was left vague.

(ii) Can his complicated theory of factorization of cyclotomic integers C_p into ideal primes be made transparent?

(iii) Can it be extended to domains other than C_p? For example, to "quadratic domains", $Z_d = \{a + b\sqrt{d} : a, b \in Z\}$, if $d \equiv 2$ or 3 (mod 4), and $Z_d = \left\{a/2 + (b/2)\sqrt{d} : \text{a and b are both even or both odd}\right\}$, if $d \equiv 1$ (mod 4)? These domains are important in the study of quadratic forms. As a rule they are *not* ufds. For instance, $Z_{-5} = \{a + b\sqrt{-5} : a, b \in Z\}$ is not. For here $6 = 2 \times 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$, where 2, 3, $1+\sqrt{-5}, 1-\sqrt{-5}$ are primes in Z_{-5}.

It was left to Dedekind to give satisfactory answers to these questions. He did this in a revolutionary work in 1871, introducing the concepts of field, ring, and ideal—in the context of the complex numbers—and formulating a broadly applicable Unique Factorization Theorem (UFT).

A central idea in this work is that of an "algebraic number field". Let a be an algebraic number—a root (recall) of a polynomial with integer coefficients—and set $Q(a) = \{q_0 + q_1a + q_2a^2 + \ldots + q_na^n : q_i \in Q\}$, Q the field of rational numbers. Dedekind showed that *all* the elements of Q(*a*) are algebraic numbers, and that Q(*a*) is a "field", called an "algebraic number field". In fact, he was the first to *define* a field.

Let now $I(a) = \{\alpha \in Q(a) :$ α is an "algebraic *integer*"$\}$; that is, α is a root of a "monic" polynomials with integer coefficients. (A polynomial is "monic" if the coefficient of the highest-degree term is 1.) Dedekind showed that I(a) is a "ring"; its elements are called "the integers of

Carl Friedrich Gauss (1777–1855)

Fermat's Last Theorem

Recall that in the seventeenth century Fermat proved FLT, the unsolvability in nonzero integers of $x^n + y^n = z^n$ $(n > 2)$, for $n = 4$. Given this result, one can readily show that it suffices to prove FLT for $n = p$, an odd prime. But over the next two centuries the theorem was proved for only three more cases: $n = 3$ (Euler, in the eighteenth century), $n = 5$ (Adrien Marie Legendre and Peter Lejeune Dirichlet, independently, in the early nineteenth century), and $n = 7$ (Gabriel Lamé (1795–1870), in 1837).

A general attack on FLT was made in 1847, again by Lamé. His idea was to factor the left side of $x^p + y^p = z^p$ into linear factors (as Euler had already done for $n = 3$, and for the Bachet equation $x^2 + 2 = y^3$) to obtain the equation $(x+y)(x+yw)(x+yw^2) \dots (x+yw^{p-1}) = z^p$, where w is a primitive p-th root of 1 $(w^p = 1, w \neq 1)$. This is an equation in the domain of "cyclotomic integers" of order p, $C_p = \{a_0 + a_1 w + a_2 w^2 + \ldots + a_{p-1} w^{p-1} : a_i \in Z\}$. Lamé now proceeded to prove FLT, using the arithmetic of C_p, as others had done before him for small values of p.

But his proof was flawed: it assumed that the arithmetical properties of Z carry over to C_p, namely that C_p is a ufd. When Lamé presented his proof to the Paris Academy of Sciences, Joseph Liouville, who was in the audience, took the floor to point out precisely that. Lamé responded that he would reconsider his proof but was confident that he could repair it. Alas, this was not to be.

Factorization of Ideals

Two months after Lamé's presentation, Liouville received a letter from Ernst Kummer, informing him that while C_p is indeed a ufd for all $p > 23$, C_{23} is not. (It was shown in 1971 that unique factorization fails in C_p for *all* $p > 23$.) But all hope was not lost, continued Kummer in his letter [12, p. 7]:

> It is possible to rescue it [unique factorization] by introducing a new kind of complex numbers, which I have called *ideal complex numbers*…. I have considered already long ago the applications of this theory to the proof of Fermat's [Last] Theorem and I succeeded in deriving the impossibility of the equation $x^n + y^n = z^n$ [for all $n > 100$].

power series. Mathematicians put their trust in such broad principles because for the most part they yielded correct results.

5.5 A Look Ahead: Foundations

Mathematicians of the seventeenth and eighteenth centuries realized that the subject they were creating was not on firm ground. For example, Newton affirmed of his fluxions that they were "rather briefly explained than narrowly demonstrated" [4, p. 201]. Leibniz said of his differentials that "it will be sufficient simply to make use of them as a tool that has advantages for the purpose of calculation" [4, p. 265]. The Berlin Academy offered a prize in 1784, hoping that "it can be explained how so many true theorems have been deduced from a contradictory supposition [namely, the existence of infinitesimals]" [6, p. 41]. Lagrange made an elaborate—but essentially misguided—response to this challenge, although his work could be justified in the contemporary setting.

In the late eighteenth and early nineteenth centuries, the work of Lagrange, Joseph Fourier, and others forced mathematicians to confront the lack of rigor in calculus. Here is Niels-Henrik Abel on the subject [11, p. 973]:

> Divergent series [employed by Newton, Euler, and others] are the invention of the devil. By using them, one may draw any conclusion he pleases, and that is why these series have produced so many fallacies and so many paradoxes.

Starting in 1821 and continuing for about half a century, a series of mathematicians, including Augustin-Louis Cauchy, Bernard Bolzano, Richard Dedekind, and Karl Weierstrass, supplied calculus with foundations, essentially as we have them today. The main features of their work were:

I. The emergence of the notion of limit as the underlying concept of calculus.
II. The recognition of the important role played—in definitions and proofs—by inequalities.
III. The acknowledgement that the validity of results in calculus must take into account questions of the domain of definition of a function. (In the eighteenth century a theorem of calculus was usually regarded as universally true by virtue of the *formal* correctness of the underlying algebra.)
IV. The realization that for a logical foundation of calculus one must have a clear understanding of the nature of the real number system, and that this understanding should be based on an *arithmetic* rather than a geometric conception of the continuum of real numbers.

The work on foundations of calculus did away "for good" with infinitesimals—used by Cauchy and his predecessors for over two centuries (two millennia, if we consider the Greek contributions). In 1960, infinitesimals were actually brought back to life, as genuine and rigorously defined mathematical objects, in the "nonstandard analysis" conceived by the mathematical logician Abraham Robinson—but that is another story!

Augustin-Louis Cauchy (1789–1857)

Problems and Projects

1. Describe some of Pascal's, Roberval's, or Wallis' work in calculus.
2. Discuss the priority dispute between Newton and Leibniz concerning the invention of calculus.
3. Write a short essay on Archimedes' *Method*.
4. Discuss Euler's use of power series.
5. Describe the essential elements in Lagrange's algebraic approach to calculus.
6. Discuss Bishop George Berkeley's critique of Newton's calculus.
7. Write an essay on the "Arithmetization of Analysis". See [1, 4, 8, 11].
8. Discuss some of the errors in calculus in the late eighteenth and early nineteenth centuries resulting from the lack of proper foundations. See [1, 5, 8].
9. Write a brief essay on the basic ideas of nonstandard analysis. See [2–4, 7].

References

1. Bottazzini, U.: The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass. Springer, New York (1986)
2. Dauben, J.W.: Abraham Robinson and nonstandard analysis: history, philosophy, and foundations of mathematics. In: Aspray, W., Kitcher, P. (eds.), History and Philosophy of Modern Mathematics, pp. 177–200. University of Minnesota Press, Minneapolis (1988)
3. Davis, M., Hersh, R.: Nonstandard analysis. Sci. Am. **226**, 78–86 (1972)
4. Edwards, C.H.: The Historical Development of the Calculus. Springer, New York (1979)
5. Gardiner, A.: Infinite Processes: Background to Analysis. Springer, New York (1979)
6. Grabiner, J.W.: The origins of cauchy's rigorous calculus. MIT Press (1981)
7. Harnik, V.: Infinitesimals from Leibniz to Robinson: time to bring them back to school. Math. Intell. **8**(2), 41–47, 63 (1986)
8. Katz, V.: A History of Mathematics, 3rd edn. Addison-Wesley, Boston (2009)
9. Kitcher, P.: The Nature of Mathematical Knowledge. Oxford University Press, New York (1983)
10. Kline, M.: Mathematical Thought from Ancient to Modern Times. Oxford University Press, Oxford (1972)
11. Kline, M.: Euler and infinite series. Math. Mag. **56**, 307–314 (1983)

Leonhard Euler (1707–1783)

methods, solved important problems, and furnished mathematicians with new ideas to help guide their researches in the decades ahead. Two central problems provided the early stimulus for these developments: reciprocity laws and FLT.

Reciprocity Laws

The "quadratic reciprocity law", the relationship between the solvability of $x^2 \equiv p \pmod q$ and $x^2 \equiv q \pmod p$, with p and q distinct odd primes, is a fundamental result, established by Gauss in 1801. A major problem, posed by him and others, was the extension of that law to higher analogues, which would describe the relationship between the solvability of $x^n \equiv p \pmod q$ and $x^n \equiv q \pmod p$ for $n > 2$. (The cases $n = 3$ and $n = 4$ give rise to what are called "cubic" and "biquadratic" reciprocity, respectively.) Gauss opined that such laws cannot even be *conjectured* within the context of the integers. As he put it: "such a theory [of higher reciprocity] demands that the domain of higher arithmetic [i.e., the domain of integers] be endlessly enlarged" [7, p. 108]. This was indeed a prophetic statement.

Gauss himself began to enlarge that domain by introducing (in 1832) what came to be known as the "gaussian integers", $Z(i) = \{a + bi : a, b \in Z\}$. He needed them to formulate a "biquadratic reciprocity law" [7]. The elements of Z(i) do indeed qualify as "integers", in the sense that they obey all the crucial arithmetic properties of the "ordinary" integers Z: They can be added, subtracted, and multiplied, and, most importantly, they obey a Fundamental Theorem of Arithmetic—every noninvertible element of Z(i) is a unique product of primes of Z(i), called "gaussian primes". The latter are those elements of Z(i) that cannot be written nontrivially as products of gaussian integers; for example, $7 + i = (2 + i)(3 - i)$, where $2 + i$ and $3 - i$ are gaussian primes [1].

A domain with a unique factorization property such as the above is called (as we have seen) a "ufd". Thus Z(i) is a ufd. Gauss also formulated a *cubic* reciprocity law, and to do that he introduced yet another domain of "integers", the "cyclotomic integers" of order 3, $C_3 = \{a + bw + cw^2 : a, b, c \in Z\}$, where, $w = (-1 + \sqrt{3}i)/2$ is a primitive cube root of 1 ($w^3 = 1, w \neq 1$). This, too, turned out to be a ufd. Higher reciprocity laws were obtained in the nineteenth and early twentieth centuries [7].

that they are the *only* (positive) solutions (see ► Section 6.4 below). Bachet's equation plays a central role in number theory to this day.

6.4 Euler and the Bachet Equation $x^2+2=y^3$

Leonhard Euler was the greatest mathematician of the eighteenth century, and one of the most eminent of all time, "the first among mathematicians", according to Lagrange. He was also the most productive ever. Although "only" four of a projected 85 or so volumes of his collected works are on number theory, they contain priceless treasures, dealing with all of the subject's existing areas and giving birth to new methods and results.

A considerable part of Euler's number-theoretic work consisted in proving Fermat's results and trying to reconstruct his methods [18]. Euler dealt with diophantine equations (among other topics in number theory) in his book *Elements of Algebra* (1770). In particular, he solved the Bachet equation $x^2+2=y^3$ by introducing a new—and most important—idea, namely factoring the equation's left-hand side. This yielded $\left(x+\sqrt{2}i\right)\cdot\left(x-\sqrt{2}i\right)=y^3$, an equation in a domain D of "complex integers", where $D=\{a+b\sqrt{2}i : a, b\in Z\}$. Euler proceeded as follows:

If a, b, and c are integers such that $ab=c^3$, and $(a, b)=1$ ((a, b) denotes the greatest common divisor of a and b), then $a=u^3$ and $b=v^3$, with u and v integers. This is a well known and easily established result in number theory. (It holds with the exponent 3 replaced by any integer, and for any number of factors a, b, ….) Euler carried it over—*without justifying the move*—to the domain D. Since, $\left(x+\sqrt{2}i\right)\left(x-\sqrt{2}i\right)=y^3$ and $(x+\sqrt{2}i, x-\sqrt{2}i)=1$ (Euler claimed, *without substantiation*, that $(m, n)=1$ in Z implies $\left(m+n\sqrt{2}i, m-n\sqrt{2}i\right)=1$ in D), it follows that $x+\sqrt{2}i=\left(a+b\sqrt{2}i\right)^3=\left(a^3-6ab^2\right)+\left(3a^2b-2b^3\right)\sqrt{2}i$ for some integers a and b. Equating real and imaginary parts we get $x=a^3-6ab^2$ and $1=3a^2b-2b^3=b(3a^2-2b^2)$. Since a and b are integers, we must have $a=\pm 1$, $b=1$, hence $x=\pm 5$, $y=3$. These, then, are the *only* solutions of $x^2+2=y^3$.

To have Euler's proof meet modern standards of rigor, one would need to define "unique factorization domain" (ufd), show that the domain D is a ufd, and verify the steps used above without justification. But Euler apparently had no compunction in viewing his solution of the diophantine equation $x^2+2=y^3$ as legitimate.

Rigor aside, Euler had taken the audacious step of introducing *complex numbers* into number theory—the study of the *positive integers*. "A momentous event had taken place", declared André Weil (1906–1998), adding: "Algebraic numbers had entered number theory—through the back door" [18, p. 242]. (An "algebraic number" is a complex number which is a root of a polynomial with integer coefficients.) While Euler had earlier wedded number theory to analysis [5, 18], he now linked number theory with algebra. This bridge-building would prove most fruitful in the following century.

6.5 Reciprocity Laws, Fermat's Last Theorem, Factorization of Ideals

Before the nineteenth century number theory consisted of many brilliant results but often lacked thematic unity and general methodology. In his masterpiece, *Disquisitiones Arithmeticae* (1801), Gauss supplied both. He systematized the subject, provided it with deep and rigorous

Further Reading

12. Andersen, K.: Cavalieri's method of indivisibles. Arch. Hist. Exact Sci. **31**, 291–367 (1985)
13. Baron, M.: The Origins of the Infi nitesimal Calculus. Dover, New York (1987)
14. Bos, H.J.M.: Differentials, higher-order differentials and the derivative in the Leibnizian calculus. Arch. Hist. Exact Sci. **14**, 1–90 (1974)
15. Boyer, C.B.: Cavalieri, limits and discarded infinitesimals. Scr. Math. **8**, 79–91 (1941)
16. Boyer, C.B.: The History of the Calculus and Its Conceptual Development. Dover, New York (1959)
17. Cajori, F.: Discussion of fluxions: From Berkeley to Woodhouse. Am. Math. Mon. **24**, 145–154 (1917)
18. Cajori, F.: Grafting of the theory of limits on the calculus of Leibniz. Am. Math. Mon. **30**, 223–234 (1923)
19. Cajori, F.: Indivisibles and "ghosts of departed quantities" in the history of mathematics. Scientia **37**, 301–306 (1925)
20. Fraser, C.: The calculus as algebraic analysis: some observations on mathematical analysis in the 18th century. Arch. Hist. Exact Sci. **39**, 317–335 (1989)
21. Grabiner, J.W.: Who gave you the epsilon? Cauchy and the origins of rigorous calculus. Am. Math. Mon. **90**, 185–194 (1983)
22. Grabiner, J.W.: The changing concept of change: the derivative from Fermat to Weierstrass. Math. Mag. **56**, 195–206 (1983)
23. Kitcher, P.: Fluxions, limits, and infinite littlenesse: a study of Newton's presentation of the calculus. Isis **64**, 33–49 (1973)

Gaussian Integers: From Arithmetic to Arithmetics

H. Grant, I. Kleiner, *Turning Points in the History of Mathematics*, Compact Textbooks in Mathematics, DOI 10.1007/978-1-4939-3264-1_6, © Springer Science+Business Media, LLC 2015

6.1 Introduction

Number theory, also known as "arithmetic", or "higher arithmetic", is the study of properties of the positive integers. It is one of the oldest branches of mathematics, and has fascinated both amateurs and professionals throughout history. Many of its results are simple to state and understand, and many are suggested by concrete examples. But results are frequently very difficult to *prove*. It is these attributes of the subject that give number theory a unique and magical charm, claimed Carl Friedrich Gauss, one of the greatest mathematicians of all time.

To deal with the many difficult number-theoretic problems, mathematicians have had to invoke—often to *invent*—advanced techniques, mainly in algebra, analysis, and geometry. So began, in the nineteenth and twentieth centuries, distinct *branches* of number theory, such as algebraic number theory, analytic number theory, transcendental number theory, and the geometry of numbers. It is in the context of algebraic number theory that we will encounter various "arithmetics".

6.2 Ancient Times

"Diophantine equations", so named after the Greek mathematician Diophantus (fl. c. 250 AD), who examined them extensively, have been a central theme in number theory. These are equations in two or more variables, with integer or rational coefficients, for which the solutions sought are integers or rational numbers. The earliest such equation, $x^2+y^2=z^2$, dates back to Babylonian times, about 1800 BC. It has been important throughout the history of number theory. Its *integer* solutions are called "Pythagorean triples".

Records of Babylonian mathematics have been preserved on clay tablets. One of the most renowned, named "Plimpton 322", consists of fifteen rows of numbers, which have been interpreted as fifteen Pythagorean triples, each triple perhaps giving the sides of a right triangle [8, p. 19]. There is no indication of how they were generated, or why (mathematics for fun?), but the listing suggests, as do other sources, that the Babylonians knew the Pythagorean theorem more than a millennium before the birth of Pythagoras (c. 570 BC).

6.3 Fermat

Pierre de Fermat was arguably the greatest mathematician of the first half of the seventeenth century—though a lawyer by profession! In mathematics he made fundamental contributions to several areas, but number theory was his special passion. In fact, he founded that subject in its modern form.

Pierre de Fermat (1601–1665)

Fermat's interest in number theory was aroused by Diophantus' acclaimed work *Arithmetica* [6]. He famously noted in the margin of Problem 8, Book II of Diophantus' book, which gave the representation of a given square as a sum of two squares, that —in contrast to that result—

> It is impossible to separate a cube into two cubes or a fourth power into two fourth powers or, in general, any power greater than the second into powers of like degree. I have discovered a truly marvelous demonstration, which this margin is too narrow to contain [4, p. 2].

Fermat thus claimed that the equation $z^n = x^n + y^n$ has no (nonzero) integer solutions if $n > 2$. This has come to be known as "Fermat's Last Theorem" (FLT), and was perhaps the most outstanding unsolved problem in number theory for 360 years. The distinguished mathematician André Weil (1906–1998) said the following about Fermat's claim [18, p. 104]:

> For a brief moment perhaps, and perhaps in his younger days, he must have deluded himself into thinking that he had the principle of a general proof [of FLT]; what he had in mind on that day can never be known.

The Princeton mathematician Andrew Wiles, who supplied a proof in 1994 [9, 14]—more than three centuries after Fermat's claim—also thought it most unlikely that Fermat had succeeded. (Fermat did not give *any* proofs in his number-theoretic work, with the exception of FLT for $n=4$, which is easier than for $n=3$.)

Another important equation considered by Fermat is the "Bachet equation", $x^2+k=y^3$ (k is an integer), named after Claude-Gaspar Bachet de Mézeriac (1581–1638), a member of an informal group of Parisian scientists. Fermat found the (positive) solutions of $x^2+2=y^3$ and $x^2+4=y^3$, namely $x=5$, $y=3$ for the first equation, and $x=2$, $y=2$ and $x=11$, $y=5$ for the second. It is easy to verify that these are solutions of the respective equations, but rather difficult to show