ARTICLES

Rigor and Proof in Mathematics: A Historical Perspective

ISRAEL KLEINER
York University
North York, Ontario, Canada M3J 1P3

> Mathematical rigor is like clothing: in its style it ought to suit the occasion, and it diminishes comfort and restricts freedom of movement if it is either too loose or too tight [**52**, p. ix].

The above observation is sound pedagogical advice. It also reflects mathematical practice and its historical evolution. Standards of rigor have changed in mathematics, and not always from less rigor to more. The notion of proof is not absolute. Mathematicians' views of what constitutes an acceptable proof have evolved. In this article we will briefly trace that evolution.

Several themes emerge:

(a) The validity of a proof is a reflection of the overall mathematical climate at any given time.

(b) The causes of transition from less rigor to more rigor (or vice versa) were, in general, not aesthetic or epistemological; there were good *mathematical* reasons for such changes.

(c) Every tightening (or relaxation) of the standards of rigor created new problems having to do with rigor.[1]

We will give a sketch of the evolution of the concept and practice of proof at its turning points—the periods that made the greatest contributions to its elucidation.[2]

1. The Babylonians

Babylonian mathematics is the most advanced and sophisticated of pre-Greek mathematics, but it lacks the concept of proof. There are no general statements in Babylonian mathematics and there is no attempt at deduction, or even at reasonable

[1] A familiar theme in mathematics: Each time a problem is solved, several new ones emerge.

[2] I am well aware that in dealing with a nearly 4000-year span of mathematical history one can hardly do justice to topics deserving more thorough treatment. On the other hand, since I am attempting to survey major trends, I touch on issues with which some readers are well acquainted. Nevertheless, I hope the overall survey will be of interest to readers. The references (they are, for ease of access, to secondary sources) should serve as entry points for further study.

explanation, of the validity of the results.[3] This mathematics deals with specific problems, and the solutions are prescriptive—do this and that and you will get the answer. The following is an example (ca. 1600 B.C.) of a typical problem and its solution [**6**, p. 69]:

> I have added the area and two-thirds of the side of my square and it is 0; 35[$\frac{35}{60}$ in sexagesimal notation]. What is the side of my square?[4]

Solution:

> You take 1, the coefficient. Two-thirds of 1 is 0; 40. Half of this, 0; 20, you multiply by 0; 20 and it [the result] 0; 6, 40 you add to 0; 35 and [the result] 0; 41, 40 has 0; 50 as its square root. The 0; 20, which you have multiplied by itself, you subtract from 0; 50, and 0; 30 is [the side of] the square.

In modern notation the problem is to solve the equation $x^2+\frac{2}{3}x=\frac{35}{60}$. The instructions for its solution can be expressed as:

$$\begin{aligned} x &= \sqrt{\left(\frac{0;40}{2}\right)^2+0;35} - \frac{0;40}{2} \\ &= \sqrt{0;6,40+0;35} - 0;20 \\ &= \sqrt{0;41,40} - 0;20 \\ &= 0;50-0;20 \\ &= 0;30. \end{aligned}$$

These instructions amount to the use of the formula

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

to solve the equation $x^2+ax=b$—a remarkable feat, indeed.

Many similar examples appear in Babylonian mathematics (see e.g. [**65**]). Indeed, the accumulation of example after example of the same type of problem indicates the existence of some form of justification of Babylonian mathematical procedures.[5] In any case, as Wilder suggests [**71**, p. 156]:

> The Babylonians had brought mathematics to a stage where two basic concepts of Greek mathematics were ready to be born—the concept of a *theorem* and the concept of a *proof*.

See [**6**], [**35**], [**65**], [**71**] for further details on this section.

[3] Mathematics without proof—a paradox?

[4] This is, of course, a "fun" problem without practical utility—mathematics for its own sake ca. 1600 B.C.! It is also noteworthy that the Babylonians are adding area to length—forbidden in the later and much more sophisticated Greek mathematics.

[5] For example, it has been suggested that the Babylonians knew the method of "completing the square" for solving quadratic equations. See [**65**].

2. Greek Axiomatics

Proof as deduction from explicitly stated postulates was, of course, conceived by the Greeks. 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 offers a foundation for mathematical reasoning was, indeed, an extraordinary development. When, how, and why this came about is open to conjecture. Various reasons—both internal and external to mathematics[6]—have been advanced for the emergence of the deductive method in ancient Greece, the so-called Greek mathematical miracle. Among the suggested reasons are:

(a) the need to resolve the "crisis" engendered by the Pythagoreans' proof of the incommensurability of the diagonal and side of the square (see [**18**]). This no doubt provided an important impetus for a critical re-evaluation of the logical foundations of mathematics.

(b) the desire to decide among contradictory results bequeathed to the Greeks by earlier civilizations (see [**65**, p. 89]). (For example, the Babylonians used the formula $3r^2$ for the area of a circle[7], the Egyptians $(\frac{8}{9} \times 2r)^2$.) This encouraged the notion of mathematical demonstration, which in time evolved into the deductive method.

(c) the nature of Greek society. Democracy in Greece required the art of argumentation and persuasion, and hence encouraged logical, deductive reasoning. Moreover, the existence of a leisure class, supported by a large slave class, was (probably) at least a necessary condition for mathematical contemplation and abstract thinking. Thus, paradoxically, both democracy and slavery apparently contributed to the emergence of the deductive method. See [**45**, Ch. 4].

(d) the predisposition of the Greeks to philosophical inquiry in which answers to ultimate questions are of prime concern. In particular, it has been argued that the axiomatic method originated in the Eleatic school of philosophy begun by Parmenides and furthered by his pupil Zeno in the early 5th century B.C. Zeno, in fact, does use the indirect method of proof in his famous paradoxes. See [**59**], but also [**37**] in which an alternate thesis is proposed.[8]

(e) the need to teach. This forced the Greek mathematicians to consider the basic principles underlying their subject. There were, in fact, about a dozen compilers of "Elements" before Euclid (see [**37**, p. 179]). It is noteworthy that the pedagogical motive in the formal organization of mathematics was also present in the works of later mathematicians (as we shall note), notably Lagrange, Cauchy, Weierstrass, and Dedekind.

The axiomatic method in Greece did not come without costs. It is paradoxical that the very perfection of classical Greek mathematics—the insistence on strict, logical

[6] Wilder [**71**] calls them "hereditary" and "environmental" stresses, respectively.

[7] There is evidence that the Babylonians also used $3\frac{1}{8}$ as an estimate for π. See [**35**, p. 11].

[8] In this connection it is interesting to note the view of A. C. Clairaut, an 18th-century mathematician and scientist, on Euclid's proofs of obvious propositions [**35**, 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.

deduction—likely contributed to its eventual decline. For this insistence precluded the use by the Greeks of such "working tools" as irrational numbers and the infinite (Eudoxus' theory of incommensurables and his method of exhaustion notwithstanding), which proved fundamental for the subsequent development of mathematics. Thus a very rigorous period in mathematics brought in its wake a long period of mathematical activity with little attention paid to rigor. Too much rigor may lead to rigor mortis.[9]

See **[6]**, **[18]**, **[35]**, **[37]**, **[59]**, **[65]**, **[70]**, **[71]** for details on this section.

3. Symbolic Notation

We take symbolism in mathematics for granted. In fact, mathematics without a well-developed symbolic notation would be inconceivable to us. We should note, however, that mathematics evolved for at least three millennia with hardly any symbols! The introduction and perfection of symbolic notation occurred largely in the sixteenth and seventeenth centuries and is due mainly to Viète, Descartes, and Leibniz. Symbolic notation proved to be the key to a very powerful method of demonstration. One need only compare Cardan's three-page derivation (in 1545) of the formula for the solution of the cubic (see **[57**, p. 63**]**) with the corresponding modern half-page proof (see **[6**, p. 311**]**). Moreover, in the absence of symbols, Cardan deals with equations with *numerical* coefficients rather than with literal coefficients that are, of course, required for a general proof.

The pedagogical advantages resulting from symbolic notation are well expressed by C. H. Edwards in his comments about Leibniz' felicitous notation for the calculus **[16**, p. 232**]**:

> It is hardly an exaggeration to say that the calculus of Leibniz brings within the range of an ordinary student problems that once required the ingenuity of an Archimedes or a Newton.

In addition to being the key to a method of demonstration and an invaluable pedagogical aid, symbolic notation also proved to be the key to a method of discovery. For example, the relation between the roots and coefficients of a polynomial equation could surely have been noticed only after symbolic notation for polynomial equations was well in place (see **[22]**). The discovery of new results was often a consequence of the intimate relation between content and form that a good notation frequently implies. For instance,

> [Leibniz'] infinitesimal calculus is the supreme example in all of science and mathematics, of a system of notation and terminology so perfectly mated with its subject as to faithfully mirror the basic logical operations and processes of that subject **[16**, p. 232**]**.[10]

[9] The predominance of rigorous thinking in Greek mathematics was, of course, not the only cause of the lack of concern for rigor during the following two millennia. See **[35]** and **[65]**.

[10] Leibniz' striving for an efficient notation for his calculus was part and parcel of his endeavor to find a "universal characteristic"—a symbolic language capable of mechanizing rational expression.

As an illustration, we cite Leibniz' discovery (and "proof") of the product rule for differentiation:

$$d(xy) = (x + dx)(y + dy) - xy = xy + x\,dy + y\,dx + dx\,dy - xy = x\,dy + y\,dx,$$

since "the quantity $dx\,dy \ldots$ is infinitely small in comparison with the rest," notes Leibniz [16, p. 255], and hence can be discarded.[11]

Euler elevated symbol-manipulation to an art. Note his uncanny derivation of the power-series expansion of $\cos x$ [22, p. 355]:
Use the binomial theorem to expand the left-hand side of the identity

$$(\cos z + i \sin z)^n = \text{cox}\, nz + i \sin nz.$$

Equate the real part to $\cos nz$ to obtain

$$\cos nz = (\cos z)^n - \frac{n(n-1)}{2!}(\cos z)^{n-2}(\sin z)^2 + \frac{n(n-1)(n-2)(n-3)}{4!}(\cos z)^{n-4}(\sin z)^4 - \cdots.$$

Now let n be an infinitely large integer and z an infinitely small number. Then

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

The above equation becomes

$$\cos nz = 1 - \frac{n^2 z^2}{2!} + \frac{n^4 z^4}{4!} - \cdots.$$

Letting $nz = x$ (Euler claims that nz is finite since n is infinitely large and z infinitely small) we finally get

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots (!).$$

This formal "algebraic analysis," so brilliantly used by Euler and practiced by most 18th-century mathematicians, 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 power series.[12]

What made mathematicians put their trust in the power of symbols? First and foremost, the use of such formal methods led to important results. A strong intuition by the leading mathematicians of the time kept errors to a minimum.[13] Moreover, the

[11] Although this derivation may seem trivial, it was only after a considerable struggle that Leibniz arrived at the correct rules for the differentiation of products and quotients.

[12] An elementary example of the use of some of these principles—descendents of Leibniz' "principle of continuity"—was the deduction, from the "identity" $1/(1+x) = 1 - x + x^2 - x^3 + \cdots$, of the equality $\frac{1}{2} = 1 - 1 + 1 - 1 + \cdots$ (obtained by setting $x = 1$). This latter result elicited mathematical, metaphysical, and theological discussion. See [35, p. 485].

[13] Errors *were* made. See, e.g., [17, p. 10] for Schwarz' counterexample to Euler's proof of the equality $f_{xy} = f_{yx}$ of the partial derivatives of a function $f(x, y)$. For more recent examples of errors made by mathematicians see [11, p. 260] and [14, p. 272].

methods were often applied to physical problems and the reasonableness of the solutions "guaranteed" the correctness of the results (and, by implication, the correctness of the methods). There was also a belief (held by Newton, among others) that mathematicians were simply uncovering God's grand mathematical design of nature.[14]

See [3], [16], [20], [22], [24], [34], [35] for further details on this section.

4. The Calculus of Cauchy

Concern about foundations was never quite absent from mathematics, but it became a dominant feature of its development in the 19th century. This century ushered in a spirit of scrutiny of the concepts and methods in various areas of mathematics, and, in particular, in analysis. This spirit is already clearly apparent in Gauss' classic *Disquisitiones Arithmeticae* of 1801.[15] Other noteworthy examples were Peacock's work in algebra and Bolzano's work in analysis. We will focus, however, on Cauchy's seminal work, begun in his *Cours d'Analyse* of 1821, of providing a rigorous foundation for the calculus.

Cauchy selected a few fundamental concepts, namely limit, continuity, convergence, derivative, and integral, established the limit concept as the one on which to base all the others, and derived by fairly modern means the major results of the calculus. That this sounds commonplace to us today is, in large part, a tribute to Cauchy's programme—a grand design, brilliantly executed. In fact, most of the above basic concepts of the calculus were either not recognized or not clearly delineated before Cauchy's time.[16]

What impelled Cauchy to make such a fundamental departure from established practice? Several reasons can be advanced.

(a) In 1784 Lagrange proposed to the Berlin Academy the foundations of the calculus as a prize problem. His lectures on the calculus at the Ecole Polytechnique were published in two influential books, in 1797 and 1799–1801. These works of Lagrange made an impact on both Bolzano and Cauchy. The methods of Lagrange and Cauchy, however, were diametrically opposed. As Lagrange put it, his books were to contain "the principal theorems of the differential calculus without the use of the infinitely small, or vanishing quantities, or limits and fluxions, and reduced to the art of algebraic analysis of finite quantities" [35, p. 430]. Thus Lagrange's foundation for the calculus was based on its reduction to algebra, for "he wanted to gain for the

[14] This belief had changed by the end of the 18th century. When Laplace gave Napoleon a copy of his *Mécanique Céleste*, Napoleon is said to have remarked [35, p. 621]: "M. Laplace, they tell me you have written this large book on the system of the universe and have never even mentioned its Creator," whereupon Laplace replied: "Sir, I have no need of this hypothesis."

[15] Even the rigor of the great Gauss was relative to his time. Thus Smale [53, p. 4] notes an "immense gap" in Gauss' proof of the Fundamental Theorem of Algebra—a gap filled only in 1920, over 100 years after Gauss "proved" the theorem.

[16] The concept of limit was only adumbrated in the 18th century. Euler defined continuity but in a sense different from Cauchy's (and ours). The differential rather than the derivative was the dominant concept in 18th-century analysis; the integral was viewed as an antiderivative. Convergence was rarely considered before the 19th century. Cauchy (along with Abel and others) "banished" divergent series—which Euler found so useful—from analysis. They began to be formally resurrected as legitimate, rigorous mathematical entities toward the end of the 19th century. See [3], [16], [20], [24], [35] for details.

calculus the certainty he believed algebra to possess" [**25**, p. 189].[17] Cauchy's aim, on the other hand, was to eliminate algebra as a basis for the calculus and thus to repudiate 18th-century practice:

> As for my methods, I have sought to give them all the rigor which is demanded in [Euclidean] geometry, in such a way as never to run back to reasons drawn from what is usually given in algebra. Reasons of this latter type, however commonly they are accepted, above all in passing from convergent to divergent series and from real to imaginary quantities, can only be considered, it seems to me, as inductions, apt enough sometimes to set forth the truth, but ill according with the exactitude of which the mathematical sciences boast. We must even note that they suggest that algebraic formulas have an unlimited generality, whereas in fact the majority of these formulas are valid only under certain conditions and for certain values of the quantities they contain [**31**, pp. 247–248].

(b) Fourier startled the mathematical community of the early 19th century with his work on what came to be known as Fourier series. Fourier claimed that:

Any function f defined over $(-l, l)$ is representable over this interval by a series of sines and cosines:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}\right),$$

where a_n, b_n are given by

$$a_n = \frac{1}{l}\int_{-l}^{l} f(t)\cos\frac{n\pi t}{l}\,dt, 2s2sb_n = \frac{1}{l}\int_{-l}^{l} f(t)\sin\frac{n\pi t}{l}\,dt.$$

Euler and Lagrange knew that *some* functions have such representations. The "principle of continuity" of 18th- and early-19th-century mathematics suggested that the above could not be true for *all* functions: Since sin and cos are continuous and periodic, the same had to be true of a sum of such terms (recall that finite and infinite sums were viewed analogously). But to refute Fourier's claim one needed—but lacked—clear notions of continuity, convergence, and the integral.[18] Cauchy rose to the challenge of clearing up the meaning of these basic concepts.

(c) Near the end of the 18th century a major social change occurred within the community of mathematicians. While in the past they were often attached to royal courts, most mathematicians since the French Revolution earned their livelihood by teaching. Cauchy was a teacher at the influential École Polytechnique in Paris,

[17] Lagrange, for example, defined the derivative of a function $y = f(x)$ as the coefficient of h in the expansion of $f(x+h)$ in a Taylor series, derived algebraically (see [**20**], [**24**] for details). In fact, to 18th-century mathematicians infinite series were part of algebra, manipulated as finite sums. Cauchy later showed that the Maclaurin expansion of

$$f(x) = \begin{cases} e^{-1/x^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

is identically zero; thus a Taylor series for $f(x)$, even if convergent, need not converge to $f(x)$ (see [**3**, p. 121]).

[18] Needless to say, Fourier's result, properly modified, was and remains one of the profound insights of analysis.

founded in 1795. It was customary at that institution for an instructor who dealt with material not in standard texts to write up notes for students on the subject of his lectures. The result, in Cauchy's case, was his *Cours d'Analyse* and two subsequent treatises. Since mathematicians presumably think through the fundamental concepts of the subject they are teaching much more carefully when writing for students than when writing for colleagues, this too might have been a contributing factor in Cauchy's careful analysis of the basic concepts underlying the calculus.

(d) The above reasons aside, it seems a "natural" process (at least from an historical perspective) that an exploratory period be followed by reflection and consolidation. Geometry in ancient Greece is a case in point. Similarly in the case of the calculus: After close to 200 years of vigorous growth with little thought given to foundations, such foundations as did exist were ripe for reevaluation and reformulation.

See [**3**], [**20**], [**22**], [**23**], [**24**], [**25**], [**31**], [**34**], [**36**] for details on this section.

5. The Calculus of Weierstrass

Cauchy's new proposals for the rigorization of the calculus generated their own problems and enticed a new generation of mathematicians to tackle them. The two major foundational problems for these successors with Cauchy's approach to the calculus were:

(a) his verbal definitions of the concepts of limit and continuity and his frequent use of the language of infinitesimals;

(b) his intuitive appeals to geometry in proving the existence of various limits.

Cauchy defines the notion of limit as follows [**31**, p. 247]:

> When the values successively attributed to the same variable approach indefinitely a fixed value, eventually differing from it by as little as one could wish, that fixed value is called the *limit* of all the others.

This is followed by a definition of infinitesimal [**31**, p. 247]:

> When the successive absolute values of a variable decrease indefinitely in such a way as to become less than any given quantity, that variable becomes what is called an *infinitesimal*. Such a variable has zero for its limit.[19]

Cauchy's definition of continuity is as follows [**3**, pp. 104–105]:

> Let $f(x)$ be a function of the variable x, and let us suppose that, for every value of x between two given limits, this function always has a unique and finite value. If, beginning from one value of x lying between these limits, we assign to the variable x an infinitely small increment α, the function itself increases by the difference $f(x+\alpha)-f(x)$, which depends simultaneously on the new variable α and on the value of x. Given this, the function $f(x)$ will be a *continuous* function of this variable within the

[19] Infinitesimals were used freely for 150 years before Cauchy's time. Cauchy, however, was the first to define them formally. Moreover, while infinitesimals were in the past usually perceived as (infinitely small) constants, Cauchy views them as variables (with zero limit). See [**44**].

> two limits assigned to the variable x if, for every value of x between these limits, the numerical value of the difference $f(x+\alpha)-f(x)$ decreases indefinitely with that of α.

These definitions suggest continuous motion—an intuitive idea. Moreover, Cauchy's formulations blur the crucial distinction between, and the placement of, the universal and existential quantifiers that precede x, ε, and δ in a modern (Weierstrassian) definition of limit and continuity. (Although Cauchy at times used $\varepsilon-\delta$ arguments in proofs of various results, he often resorted instead to the language of infinitesimals.) These shortcomings were the source of two major errors: Cauchy failed to distinguish between pointwise and uniform continuity of a function and between pointwise and uniform convergence of an infinite series of functions. Thus Cauchy "proved" that a convergent series of continuous functions is a continuous function. The proof, in which Cauchy uses infinitesimals freely, goes as follows:

Let

$$s(x)=\sum_{i=1}^{\infty}u_i(x),\qquad s_n(x)=\sum_{i=1}^{n}u_i(x),\qquad r_n(x)=\sum_{i=n+1}^{\infty}u_i(x),$$

and let α be an infinitesimal. Then $s(x+\alpha)-s(x)=[s_n(x+\alpha)-s_n(x)]+[r_n(x+\alpha)-r_n(x)]$. Since $u_i(x)$ are continuous, $u_i(x+\alpha)-u_i(x)$ is infinitesimal, hence so is $s_n(x+\alpha)-s_n(x)$ (being a finite sum of such terms). Since $\sum_1^\infty u_i(x)$ converges, $r_n(x)$ is infinitesimal for sufficiently large n; the same holds for $r_n(x+\alpha)$. Hence $r_n(x+\alpha)-r_n(x)$ is infinitesimal and thus so is $s(x+\alpha)-s(x)$. Thus an infinitesimal increment in x produces an infinitesimal increment in $s(x)$, hence $s(x)$ is continuous.

The use of infinitesimals in the proof masks the distinction between

$$(\forall\varepsilon)(\forall x)(\exists N)\left(\left|\sum_{N+1}^{\infty}u_i(x)\right|<\varepsilon\right)\quad\text{and}\quad(\forall\varepsilon)(\exists N)(\forall x)\left(\left|\sum_{N+1}^{\infty}u_i(x)\right|<\varepsilon\right),$$

and thus the distinction between pointwise and uniform convergence of the series $\sum_1^\infty u_i(x)$. See [3, p. 110] or [31, p. 254] for further details.

The above result is of course false; this was first pointed out in 1826 by Abel, who showed that the series $\sin x-\sin 2x/2+\sin 3x/3-\cdots$ converges to a function discontinuous at $x=(2n+1)\pi$ for all integers n (see [3, p. 113]). It took another 20 years, however, to determine where Cauchy went wrong![20] One was dealing with subtle concepts indeed.

Other counterexamples to plausible and widely held notions appeared during the half century following Cauchy's publication of his *Cours* and *Résumé*. Among the most unexpected was Weierstrass' example of a continuous nowhere-differentiable function $f(x)=\sum_{n=1}^{\infty}b^n\cos(a^n\pi x)$, a an odd integer, b a real number in $(0,1)$, and $ab>1+3\pi/2$. Cauchy and his contemporaries believed (and "proved") that a continuous function is differentiable except possibly at isolated points![21] See [34].

Since Cauchy's definitions of the fundamental concepts of the calculus were given in terms of limits, proofs of the existence of limits of various sequences and functions were of crucial importance. Thus Cauchy's solutions to the 18th century's lack of rigor

[20] Lakatos [40, p. 127] argues that it is a false reading of history to view Cauchy's proof as erroneous (see also p. 311). In [41] he gives a reconstruction in terms of Robinson's non-standard analysis of Cauchy's arguments. See also [44].

[21] Given the mathematicians' prevailing geometric conception of continuity (see below) and their notions of function (see [34]), this "result" is not surprising.

generated new problems. Having formulated the basic notions of the calculus algebraically, Cauchy now resorted to intuitive geometric arguments to establish a number of the fundamental results of analysis. For example, he claimed [**31**, p. 261] that

> A remarkable property of continuous functions of a single variable is to be able to be represented geometrically by means of straight lines or continuous curves,

and used this "remarkable property" of continuous functions (which, given our conceptions of function and continuity, is, of course, incorrect, as Weierstrass showed later with his counterexample) to give a (necessarily intuitive) geometric proof of the Intermediate Value Theorem.[22] Other (correct) results that Cauchy accepted on intuitive grounds are that an increasing sequence bounded from above has a limit, and that a (so-called) Cauchy sequence converges. Cauchy used these results to establish, among other things, the existence of the integral of a continuous function, and to give (in an appendix to the *Cours*) an analytic proof of the Intermediate Value Theorem. See [**16**, pp. 311, 318], [**22**, pp. 167, 170], and [**31**, p. 261].

Weierstrass and Dedekind, among others, determined to remedy this "mixture of algebraic formulation and geometric justification which Cauchy favored [and which] did not provide full comprehension of the major results of function theory" [**31**, p. 264]. Dedekind's expression of the prevailing state of affairs is revealing [**13**, pp. 1–2]:

> As professor in the Polytechnic School in Zürich I found myself for the first time obliged to lecture upon the elements of the differential calculus and felt more keenly than ever before the lack of a really scientific foundation for arithmetic. In discussing the notion of the approach of a variable magnitude to a fixed limiting value, and especially in proving the theorem that every magnitude which grows continually, but not beyond all limits, must certainly approach a limiting value, I had recourse to geometric evidences. Even now such resort to geometric intuition in a first presentation of the differential calculus, I regard as exceedingly useful, from the didactic standpoint, and indeed indispensable if one does not wish to lose too much time. But that this form of introduction into the differential calculus can make no claim to being scientific, no one will deny. For myself this feeling of dissatisfaction was so overpowering that I made the fixed resolve to keep meditating on the question till I should find a purely arithmetic and perfectly rigorous foundation for the principles of infinitesimal analysis. The statement is so frequently made that the differential calculus deals with continuous magnitude and yet an explanation of this continuity is nowhere given; even the most rigorous expositions of the differential calculus do not base their proofs upon continuity but, with more or less consciousness of the fact, they either appeal to geometric notions or those suggested by geometry, or depend upon theorems which are never established in a purely arithmetic manner. Among these, for example, belongs the above-mentioned theorem, and a more careful investigation convinced me that this theorem, or any one equivalent to it, can be regarded in some way as a sufficient basis for infinitesimal analysis. It then only remained to discover its true origin in

[22] The proof amounted to noting that if $f(a)$ and $f(b)$ differ in sign then the graph of f must cross the x-axis, hence $f(c)=0$ for some $c \in (a,b)$. See [31, p. 261].

> the elements of arithmetic and thus at the same time to secure a real definition of the essence of continuity.

Establishing theorems in a "purely arithmetic manner" implied what came to be known as the "arithmetization of analysis." Since the inception of the calculus (and even in Cauchy's time) the real numbers were viewed geometrically, without explicit formulation of their properties. Since the real numbers are in the foreground or background of much of analysis, proofs of theorems were of necessity intuitive and geometric. Dedekind's and Weierstrass' astute insight recognized that a rigorous, arithmetical definition of the real numbers would resolve the major obstacle in supplying a rigorous foundation for the calculus.[23]

The other remaining task was to give a precise "algebraic" definition of the limit concept to replace Cauchy's intuitive, "kinematic" conception. This was accomplished by Weierstrass when he gave his "static" definition of limit in terms of inequalities involving ε's and δ's—the definition we use today (at least in our formal, rigorous incarnation).[24] Thus Weierstrass also did away with infinitesimals, which were used freely by Cauchy and his predecessors for about two centuries.[25]

Looking back at 2500 years of the evolution of the notions of rigor and proof, we note that not only have the standards of rigor changed, but so have the mathematical tools used to establish rigor. Thus in ancient Greece, a theorem was not properly established until it was geometrized. In the Middle Ages and the Renaissance, geometry continued to be the final arbiter of mathematical rigor (even in algebra). Mathematicians' intuition of space appeared, presumably, more trustworthy than their insight into number—a continuing legacy of the consequences of the "crisis of incommensurability" in ancient Greece. The calculus of the 17th and especially the 18th century was no longer easily justifiable in geometric terms, and algebra became the major tool of justification (such as there was). There was a mix of the algebraic and geometric in Cauchy's work. With Weierstrass and Dedekind in the latter part of the 19th century, arithmetic rather than geometry or algebra had become the language of rigorous mathematics. To Plato, God ever geometrized, while to Jacobi, He ever arithmetized.[26] The logical supremacy of arithmetic, however, was not lasting. In the 1880s Dedekind and Frege undertook a reconstruction of arithmetic based on ideas from set theory and logic.[27] The ramifications of this event will be considered below.

See [3], [16], [26], [31], [35] for details on this section.

[23] It is noteworthy that both Weierstrass and Dedekind presented their ideas on the rigorization of the calculus in *lectures* at universities. As in Cauchy's case, so here too pedagogical considerations seemed to have been a motive in the search for careful, rigorous formulations of basic mathematical concepts.

[24] It may seem ironic that inequalities, used in the 18th century for estimation, and ε, used by some to indicate error, became in the hands of Weierstrass the very tools of precision.

[25] The story of infinitesimals is similar to that of divergent series (see footnote 16): About a century after Weierstrass had banished infinitesimals "for good" (so we all thought until 1960), they were brought back to life by Abraham Robinson as genuine and rigorously defined mathematical objects.

[26] The creation of non-Euclidean geometry, and the appearance of geometrically nonintuitive examples such as continuous nowhere-differentiable functions must have accelerated this dethroning of geometry.

[27] Sets entered mathematics considerably earlier than the 1880s, namely in connection with the arithmetization of analysis in the 1860s, in the sense that the various definitions of real numbers used (implicitly or explicitly) infinite sets of rationals as completed entities. This novel idea, soon to be elaborated by Cantor into a full-fledged theory, aroused considerable controversy and, subsequently, genuine foundational problems. Thus having resolved one important foundational problem, Weierstrass, Dedekind, et al. introduced another.

6. The Reemergence of the Axiomatic Method

Our emphasis on analysis in the last two sections is due to the fact that the most important strides in the area of rigor in the 19th century were made in analysis. However, algebra, arithmetic, and geometry were also being given careful scrutiny during this period. Moreover, mathematical logic came into being in 1847 with Boole's *The Mathematical Analysis of Logic*. All this led to a rebirth of the axiomatic method late in the 19th century. We describe these developments very briefly.

The abstract concept of a group arose from different sources. Thus polynomial theory gave rise to groups of permutations, number theory to groups of numbers and of "forms" (nth roots of unity, integers mod n, equivalence classes of binary quadratic forms), and geometry and analysis to groups of transformations. Common features of these concrete examples of groups began to be noted, and this resulted in the emergence of the abstract concept of a group in the last decades of the 19th century (see **[32]** for details). Similar observations apply to the emergence of the concepts of ring (see **[33]**), field, and (to a lesser extent) vector space.

The arithmetization of analysis reduced the foundations of the subject to that of real numbers. These were defined in terms of rational numbers. The reduction of the rationals to the positive integers soon followed.[28] There remained the problem of the foundations of the positive integers (i.e., arithmetic). This was addressed in different ways by Dedekind, Peano, and Frege during the last two decades of the 19th century. All three, however, used axiom systems to define the positive integers (see **[13]**, **[35]**).

One of the consequences of the creation of non-Euclidean geometry was a reexamination of the foundations of Euclidean geometry and, more broadly, of axiomatic systems in general. Pasch, Peano, and Hilbert pioneered the development of the modern axiomatic method (late in the 19th century) through a careful analysis of the foundations of geometry (see **[35]**, **[70]**).

Boole, by virtue of his work in mathematical logic and in (what we call today) Boolean algebra, was among the first to promote the view of the arbitrary nature of axioms allowing for different interpretations. In *The Mathematical Analysis of Logic*, Boole subscribes to what was at that time a very novel point of view [**70**, p. 116]:

> The validity of the processes of analysis does not depend upon the interpretation of the symbols which are employed, but solely upon the laws of their combination. Every system of interpretation which does not affect the truth of the relations supposed, is equally admissible.

The rise of the axiomatic method was gradual and slow (see e.g. [**32**, p. 207]). By the early 20th century, however, the axiomatic method was well established in a number of major areas of mathematics.

In algebra there were major works in group theory (1904), field theory (1910), and ring theory (1914), crowned by Emmy Noether's groundbreaking papers of the 1920s. In analysis there were Fréchet's thesis of 1906 on function spaces (in which a definition of metric space appears), E. H. Moore's work (of the same year) on "general analysis" (an axiomatic formulation of features common to linear integral equations and infinite systems of linear algebraic equations), Banach's researches on

[28] Note that the historical evolution of the logical foundations of the number system—from the reals to the rationals to the integers—is the reverse of the sequence usually presented in textbooks.

Banach spaces (1922), and von Neumann's axiomatization of Hilbert space (1929).[29] In topology Hausdorff defined a topological space in terms of neighborhoods (1914) and P. S. Alexandroff began to develop homology theory (1928) following conversations with E. Noether. In geometry Hilbert's *Foundations of Geometry* in 1899 was most influential; Veblen and Young's two-volume abstract treatment of projective geometry (1910–1919) also made a significant impact. In set theory there was Zermelo's axiomatization of set theory in 1908, followed by Fraenkel's improvements in 1921 and von Neumann's version in 1925; and, finally, in mathematical logic there was Russell and Whitehead's prodigious three-volume *Principia Mathematica* (1910–1913). See [4], [32], [33], [35], [70] for details of the above.

The axiomatic method, surely one of the most distinctive features of 20th-century mathematics, truly flourished in the early decades of the century. Bourbaki, among its most able practitioners and promoters, gives an eloquent description of the essence of the axiomatic method at what was perhaps the height of its power (in 1950):

> What the axiomatic method sets as its essential aim, is exactly that which logical formalism by itself can not 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 [4, p. 223].

In this article ([4]) Bourbaki presents a panoramic view of mathematics organized around (what he calls) "mother structures"—algebraic, ordered, and topological structures, and various substructures and cross-fertilizing structures. It must have been an alluring, even bewitching, view of mathematics to those growing up (mathematically) during this period.

There are significant differences between Euclid's axiomatics and their modern incarnation in the last decades of the 19th century and the early decades of the 20th century. Euclid's axioms are idealizations of a concrete physical reality and are thus viewed as self-evident truths—a Platonic view, describing a pre-existing reality. In the modern view axioms are neither self-evident nor true—they are simply assumptions about the relations among the undefined (primitive) terms of the axiomatic system.[30] Thus in a modern axiom system the axioms, and hence also the theorems, are *devoid of meaning*. Moreover, such an axiomatic system need not be categorical; that is, it may admit of essentially different (nonisomorphic) interpretations (models),

[29] When Von Neumann was invited to Göttingen in the 1920s to speak about linear operators on Hilbert space, Hilbert, who was in the audience, is reported to have asked: "Yes, Herr Von Neumann, but what actually is a Hilbert space?" The new developments in axiomatics at this time began to overtake even the great Hilbert.

[30] Apparently as early as 1891 Hilbert highlighted this point 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" [69, p. 14].

all of which satisfy the same axioms—a fundamentally novel idea. The modern axiomatic method is thus a unifying and abstracting device. Moreover, while the chief role played by the axiomatic method in ancient Greece was (probably) that of providing a consistent foundation, it became in the first half of the 20th century also a tool of research. In addition, the axiomatic method was at times indispensable in clarifying the status of various mathematical methods and results (e.g., the axiom of choice, the continuum hypothesis) when the mathematicians' intuition provided little guide. The method also came to be the arbiter of rigor and precision in mathematics (and beyond).[31] Thus the sometimes opposed activities of discovery and demonstration coexisted within the axiomatic method.[32]

The modern axiomatic method was not an unmitigated blessing, however (as we shall see). Although some (e.g., Hilbert) claimed that it is the central method of mathematical thought, others (e.g., Klein) argued that as a method of discovery it tends to stifle creativity. And it has its limitations as a method of demonstration.

See **[4]**, **[12]**, **[18]**, **[32]**, **[33]**, **[35]**, **[69]**, **[70]** for further details.

7. Foundational Issues

We are referring here to the three philosophies of mathematics—formalism, logicism, and intuitionism—which arose in the first decades of the 20th century and that dealt with the nature, meaning, and methods of mathematics, and thus, in particular, with questions of rigor and proof in mathematics. Although, as noted, these were 20th-century developments, they had deep roots in the mathematics of the 19th century.

The 19th 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 16th and 17th 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. In the 19th century, however, the introduction of non-Euclidean 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. Cantor's dictum that "the essence of mathematics lies in its freedom" became a reality—but one to which many mathematicians took strong exception, as the following quotations indicate.

> 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 **[33**, p. 233].

[31] This was also the case, of course, in ancient Greece. At the same time, there is perhaps no better way to bring out the differences between Greek and modern axiomatics than to compare Euclid's *Elements* with Hilbert's *Foundations of Geometry*. The comparison makes it starkly clear how standards of rigor have evolved.

[32] For example, Gray **[27**, p. 182] notes that Desarguean and non-Desarguean geometries "could never have been discovered without [the axiomatic] method".

The reservations are John Graves', who communicated them to his friend Hamilton in 1844, shortly after the latter had invented the quaternions. The "supernatural properties" referred mainly 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? [35, p. 1198]

This was Kronecker's damning praise of Lindemann, who proved in 1882 that π is transcendental (and 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 [35, 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 [35, p. 973].
>
> 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 [35, p. 1035].

The above quotations, from Hermite (in 1893), Poincaré (in 1899), and again Hermite (in 1905), respectively, are a reaction to various examples of "pathological" functions given 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 (see [34]).

> Later generations will regard *Mengenlehre* [Set Theory] as a disease from which one has recovered [35, p. 1003].

This is Poincaré again, speaking (in 1908) about Cantor's creation of set theory, in particular in connection with the paradoxes that had arisen in the theory.[33]

The above sentiments, expressed by some of the leading mathematicians of the period, are suggestive of the impending crisis. Although mathematical controversies had arisen before the 19th century (e.g., the vibrating-string controversy between D'Alembert and Euler), these were isolated cases. The frequency and intensity of the disaffection expressed in the 19th century was unprecedented and could no longer be ignored. The result was a split among mathematicians concerning the way they viewed their subject. Its formal expression was the rise in the early 20th century of three schools of mathematical thought, three philosophies of mathematics—logicism, formalism, and intuitionism. This was the first *formal* expression by mathematicians of what mathematics is about and, in particular, of what proof in mathematics is about.[34] The notion of proof—its scope and limits—became a subject of study *by mathematicians*.

[33] Compare Poincaré's position with that of Hilbert, the other giant of this period: "No one shall expel us from the paradise which Cantor created for us" [35, p. 1003].

[34] The "crises" in ancient Greece following Zeno's paradoxes and the proofs of incommensurability might have given rise to similar debates and subsequent formal resolutions, but we have little evidence of that.

The logicist thesis, expounded in the monumental *Principia Mathematica* of Russell and Whitehead, advocated that mathematics is part of logic. Mathematical concepts are expressible in terms of logical concepts; mathematical theorems are tautologies (i.e., 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 Frege on mathematical logic and the foundations of arithmetic, and by the espousal of mathematical logic by 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 (e.g., it is only on p. 362 of the *Principia* that Russell and Whitehead show that $1 + 1 = 2$!—see [**12**, p. 334]).[35] There were, moreover, serious technical difficulties in the implementation of the logicist thesis (see [**36**], [**68**]).

The most serious debate within the mathematical community—still unresolved—goes on between the adherents of the formalist and intuitionist schools. The formalist thesis, with Hilbert as its main exponent, entails viewing 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 (1920s), the axiomatic method had (as we noted) embraced much of algebra, arithmetic, analysis, set theory, and mathematical logic. Even though Zermelo's axiomatization of set theory in 1908 seemed to have avoided the paradoxes of set theory, there was no assurance that they would not reemerge in one form or another. 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.[36]

The formalists have been accused of removing all meaning from mathematics and reducing it to symbol manipulation. The charge is unfair. Hilbert's aim was to deal with the *foundations* of mathematics rather than with the daily practice of the mathematician. (The same can, of course, be said of Russell and Whitehead's objective in connection with the logicist thesis.) And to show that mathematics is free of inconsistencies one first needed to formalize the subject. It was formalism in the service of informality.

As we know, Hilbert's grand design was laid to rest by Gödel's incompleteness theorems of 1931. These showed the inherent limitations of the axiomatic method: The consistency of a large class of axiomatic systems (including those for arithmetic and set theory) cannot be established within the systems. Moreover, if consistent, these systems are incomplete (see [**12**], [**36**], [**66**] for details).[37] Chaitin notes [**8**, p. 51] that Gödel's work "demands the surprising and, for many, discomforting conclusion

[35] "If the mathematical process were really one of strict, logical progression," observe De Millo et al. [**14**, p. 272], "we would still be counting on our fingers."

[36] These methods came to be known as "metamathematics" or "proof theory." For recent developments in proof theory see [**19**].

[37] In connection with the first result, Weyl remarked: "God exists since mathematics is consistent and the devil exists since we cannot prove the consistency" [**35**, p. 1206]. The second result has elicited the comment that Gödel gave a formal demonstration of the inadequacy of formal demonstrations.

that there can be no definitive answer to the question "What is a proof?" Just as in the 19th century, following the invention of non-Euclidean geometries, noncommutative algebras, and other developments, mathematics lost its claim to (absolute) truth, so in the 20th century, following Gödel's work, it lost its claim to certainty.[38] (Although Gödel's results are of fundamental philosophical consequence, they have not affected the daily work of most mathematicians.)[39]

The intuitionists, headed by L. E. J. Brouwer, claimed that no formal analysis of axiomatic systems is necessary. In fact, mathematics should not be founded on systems of axioms. The mathematician's intuition, beginning with that of number, will guide him in avoiding contradictions. He 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.[40]

Among the results unacceptable to the intuitionists is the law of trichotomy: Given any real number N, either $N > 0$ or $N = 0$ or $N < 0$. Brouwer gave the following example to substantiate the point [**12**, p. 369]:

Define a real number $\hat{\pi}$ as follows:

(a) $\hat{\pi} = \pi$ if π does not have 100 successive zeros in its decimal expansion. If π has 100 successive zeros in its decimal expansion then

(b) $\hat{\pi} = 3.a_1a_2 \dots a_{n-1}$, where $3, a_1, a_2, \dots, a_{n-1}$ are the first n digits of π followed by 100 successive zeros, if n is odd.

(c) $\hat{\pi} = 3.b_1b_2 \dots b_{n-1}1$, where $3, b_1, b_2, \dots, b_{n-1}$ are the first n digits of π followed by 100 successive zeros, if n is even.

Let $N = \hat{\pi} - \pi$. Clearly $N = 0$, $N < 0$, or $N > 0$ if (a), (b) or (c), respectively, occurs. But we are unable, Brouwer argues, to determine which of (a), (b), or (c) occurs, hence we cannot decide which of $N = 0$, $N < 0$, or $N > 0$ holds. Thus the law of trichotomy fails.

The construction of $\hat{\pi}$ from π can be repeated, with like conclusions, to obtain $\hat{u}$ from any irrational number u. Also, "100 successive zeros" can be replaced by "10^k successive zeros (any k)." Thus even if the question of which of (a), (b), or (c) above occurs will some day be settled (for π), there are infinitely (uncountably) many similar questions, not all of which can be settled "constructively".

A prominent feature of 19th-century mathematics was nonconstructive existence results. (These were almost unknown before the 19th century.) Thus, Gauss' fundamental theorem of algebra proved the existence of roots of a polynomial equation without showing how to find them. 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 often was 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

[38] The notion that absolute truth can be attained in mathematics goes back to Descartes and Leibniz in the 17th century (see [**28**]). In the 19th century truth in mathematics was replaced by validity (relative truth) and, in the 20th century, certainty by faith. (For a formal, 20th-century notion of truth in mathematics and its relation to proof see [**60**].)

[39] See, however, [**8**] for a discussion of a connection between Gödel's theorems and random numbers.

[40] 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" [**36**, p. 246].

polynomial ring. Dedekind constructed the real numbers by using completed infinities. Such examples abound. All were rejected by the intuitionists.[41] On the other hand, the proofs of the intuitionists are certainly acceptable to the formalists.[42] Manin thus suggests that the mathematician "should at least be willing to admit that proof can have objectively different 'degrees of proofness' " [47, p. 17]. See [7], [36], [46] for details.

The differences between the formalists and the intuitionists (and their 19th-century forerunners) 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—J. von Neumann and H. Weyl, respectively—indicate:

> 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! [67, 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 [68, p. 13].

It is probably safe to say, however, that most mathematicians are untroubled, at least in their daily work, about the debates concerning the various philosophies of mathematics. Davis and Hersh [12, p. 318] put the issue in perspective:

> 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.[43]

Weyl puts 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

[41] Weyl said of nonconstructive proofs that they inform the world that a treasure exists without disclosing its location [35, p. 1203].

[42] Many results in analysis, and more recently in algebra, have been reconstructed, thanks to the pioneering effort of Errett Bishop, using finitistic methods. (See [2], [5], [46].) In fact, as early as 1924 Brouwer and Weyl gave constructive proofs yielding a root of a complex polynomial. But of what use is a constructive root if it may take up to 10^{10} years to find it!

[43] Another point of philosophical contention is between Platonists, who believe that mathematics is discovered, and formalists, who claim that it is invented (see [12], [36] for details). Davis and Hersh suggest that "the typical working mathematician is a Platonist on weekdays and a formalist on Sundays" [12, p. 321].

> language or music, of primary originality, whose historical decisions defy complete objective rationalization [**36**, p. 6].

For elaboration of various points discussed in this section see [**2**], [**5**], [**7**], [**8**], [**12**], [**19**], [**21**], [**30**], [**36**], [**46**], [**66**], [**67**], [**68**], [**69**].

8. The Age of the Computer

It may be presumptuous (if not foolhardy) to speak of mathematical trends in the last third of the 20th century. However, to a first approximation, while mathematics in the century's first two thirds (especially in the period 1930–1960) stressed the formulation of general methods and abstract theories (e.g., abstract algebra, algebraic topology, the theory of distributions, homological algebra, category theory), more attention has since been paid to the solution of specific problems (e.g. the four-color problem, the Bieberbach conjecture, Mordell's conjecture, the Poincaré conjectures).[44] The computer, no doubt, played a role in this development. It has helped stimulate the growth of new mathematical fields (e.g. algebraic coding theory, theory of automata, analysis of algorithms, optimization theory) and has aided in the revival of older fields (e.g. combinatorics, graph theory). It has also assisted in making, testing, and disproving conjectures and, more recently, in proving theorems.[45] Neither the axiomatic method nor strict adherence to very rigorous mathematical proof are hallmarks of these developments. These changes have occasioned a rethinking of the meaning and role of proof in mathematics. The catalyst has been Appel and Haken's 1976 computer-aided proof of the four-color theorem.[46] The proof required the verification, by computer, of 1,482 distinct configurations. Some critics argued that this type of proof was a major departure from the traditional mathematical proof. They advanced several reasons:

(a) The proof contained thousands of pages of computer programs *that were not published* and were thus not open to the traditional procedures of verification by the mathematical community. The proof was "not surveyable", in the words of Tymoczko, one of its forceful critics (see [**62**] and responses in [**15**] and [**58**]), and was thus "*permanently and in principle* incomplete" [**12**, p. 380].

(b) Both computer hardware and computer software are subject to error. Hence also the tendency to feel that verification of the computer results by independent computer programs was not as reliable as the standard method of checking proofs. This introduces a measure of quasi-empiricism into the proof of the four-color theorem—the computer is an experimental tool.

(c) "Proof, in its best instances, increases understanding by revealing the heart of the matter" note Davis and Hersh [**12**, p. 151]. "A good proof is one which makes us wiser", echoes Yu. I. Manin [**47**, p. 18]. Thus, even if we believe that the proof of the

[44] Clearly many counterexamples to this trend can be given; and, of course, the general theories were instrumental in the solution of these major problems.

[45] "The intruder [the computer] has changed the ecosystem of mathematics, profoundly and permanently," asserts Lynn Steen [**55**, p. 34].

[46] It is not the only instance of computer-assisted proofs. They have been used to test for primes, to verify for various values of p the Fermat conjecture that $x^p + y^p = z^p$ has no nontrivial integer solutions for $p > 2$, and recently also in functional analysis (see [**50**] and [**55**]). Most recently (December 1988) they have been employed to help prove the nonexistence of finite projective planes of order 10 (see [**10**]). It is safe to say that computer-aided proofs are here to stay.

four-color theorem is valid, we cannot *understand* the theorem unless we are (or can be) involved in the *entire* process of proof; and that is not possible in this case except for the very few.

The objections to the proof of the four-color theorem apply, *mutatis mutandis*, to the proofs of at least two other major theorems. The first one is the proof by Feit and Thompson (in the 1960s) of the solvability of all finite groups of odd order, and the other is the classification, carried out jointly by many mathematicians (in the 1980s), of finite simple groups. The first proof takes up over 300 pages of an entire issue of the *Pacific Journal of Mathematics* and is based on much previous work.[47] The second proof consists of over 11,000 pages(!) of close mathematical reasoning scattered in many journals over many years. Daniel Gorenstein, one of the major contributors to the field, said of the proof [**15**, pp. 811–812]:

> ...it seems beyond human capacity to present a closely reasoned, several-hundred-page argument with absolute accuracy...how can one guarantee that the "sieve" has not let slip a configuration which leads to yet another simple group? Unfortunately, there are no guarantees—one must live with this reality.

Speaking of the Feit-Thompson Theorem (and others whose proofs are very long), Jean-Pierre Serre observes [**9**, 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.[48]

Serre continues:

> I am also uneasy with some topics, mainly in differential topology, where the author draws a complicated picture (in two dimensions), and asks you to accept it as a proof of something taking place in five dimensions or more. Only the experts can "see" whether such a proof is correct or not—if you can call this a proof.

Largely as a result of these developments, a novel philosophy of mathematical proof seems to be emerging. It goes under various names—public proof, quasi-empiricist proof, proof as a social process. Its essence, according to its advocates, is that *proofs are not infallible*. Thus, mathematical theorems cannot be guaranteed absolute certainty.[49] And this applies not only to the theorems requiring very long proofs or the assistance of a computer, but to many "run of the mill" theorems. This is so because proofs of theorems usually rely on the correctness of other theorems. And published

[47] Chevalley once undertook to give a complete account of the proof in a seminar, but gave up after two years (see [**9**, p. 11]).

[48] There are other examples of very long proofs—e.g. the proofs of the two Burnside conjectures (ca. 500 pages apiece)—[**47**, p. 17]. See also [**14**], [**38**], [**49**], [**54**]). Some believe (see [**38**]) that long proofs are becoming the norm rather than the exception; the reason is that there are, in their view, relatively few interesting results with short proofs compared to the total number of interesting mathematical results. On the other hand, Joel Spencer suggests that the mathematical counterpart of Einstein's credo that "God does not play dice with the universe" is that "short interesting theorems have short proofs" [**54**, p. 366]. The four-color theorem and several others are currently counterexamples to this claim.

[49] It is an uncertainty quite distinct from that enunciated in Gödel's theorem.

proofs, it is argued, are usually read carefully only by the author (and perhaps by some referees) and thus mistakes are inevitable:

> Stanislaw Ulam estimates that mathematicians publish 200,000 theorems every year. A number of these are subsequently contradicted or otherwise disallowed, others are thrown into doubt, and most are ignored. Only a tiny fraction come to be understood and believed by any sizable group of mathematicians [**14**, p. 272].

The truth of a theorem, then, has a certain probability, usually < 1, attached to it. The probability increases as more mathematicians read, discuss, and use the theorem. In the final analysis, the acceptance of a theorem (i.e., the acceptance of the validity of its proof) is a social process and is based on the confidence of the mathematical community in the social systems that it has established for purposes of validation:[50]

> If a theorem has been published in a respected journal, if the name of the author is familiar, if the theorem has been quoted and used by other mathematicians, then it is considered established [**12**, p. 390].

Imre Lakatos, in a brilliant polemic [**40**], also comes to the conclusion that mathematics is fallible, although his focus and arguments differ from those in the above analysis. Mathematical theorems, Lakatos claims, are not immutable—they are subject to constant examination and possible rejection through counterexamples. Proofs are not instruments of justification but tools of discovery, to be employed in the development of concepts and the refinement of conjectures. The interplay between conjecture, proof, counterexample, and refinement of conjecture is the lifeblood of mathematics. For instance, a counterexample may compel us to tighten a definition or to broaden a theorem. These ideas are masterfully illustrated with the example of the history of the Descartes-Euler formula $V - E + F = 2$ for a polyhedron. A proof is first presented, then counterexamples are introduced, the conjecture $V - E + F = 2$ is refined (i.e., the notion of polyhedron is refined), and a new proof is given. The "give-and-take" of this historical-philosophical-pedagogical interplay encompasses about 200 years of historical analysis and continues (in [**40**]) for over 100 pages.[51]

Finally, there has recently been another interesting development in the evolution of the *concept* of proof. It has to do with the notion of *probabilistic proofs*. It has been shown that some results, even if theoretically decidable, have such long proofs that they can never be written down—neither by humans nor by computer. This is the case, for example, of almost all the familiar decidable results in logic (see [**14**],

[50] Wilder's ideas about the cultural basis of mathematics, although predating the current debate, are also relevant to the issues discussed here. See [**71**].

[51] Examples of the interplay between theorem, proof, and counterexample abound. In ancient times the Pythagorean theory of proportion applied only to commensurable magnitudes until the "counterexample" of the incommensurability of the side and diagonal of a square was discovered; a new concept of ratio was then introduced and the theory of proportion was revised (see [**65**]). In more recent times, Cauchy "proved", as we indicated earlier, that the sum of an infinite series of continuous functions is continuous; following Abel's counterexample, the concept of uniform convergence was introduced and the above result and its proof were revised. See [**31**, Ch. 10] or [**40**, Appendix 1] for details.

[48], **[56]**) as well as of tests of large numbers for primality. Michael Rabin proposed (in 1976) to relax the notion of proof by allowing probabilistic proofs (see **[51]**). For example, he found a quick way to determine, with a very small probability of error (say one in a billion), whether or not an arbitrarily chosen large number is a prime.[52] (Thus he has shown that $2^{400} - 593$ is a prime "for all practical purposes.")[53] Another instance of a probabilistic proof comes from graph theory. If two graphs are nonisomorphic, it is very difficult to establish this rigorously, but easy to show it with very high probability.

Some have argued that there is no essential difference between such probabilistic proofs and the deterministic proofs of standard mathematical practice. Both are convincing arguments. Both are to be believed with a certain probability of error. In fact, many deterministic proofs, it is claimed, have a higher probability of error than probabilistic ones. The counter argument is that there is a fundamental *qualitative* difference between the two types of proof. Although both may be subject to error, an important philosophical distinction must be made. If probabilistic proofs were routinely admitted into the domain of mathematics, this would considerably strengthen the thesis of the quasi-empirical nature of mathematics and would entail a radical departure from the traditional view of mathematics. The debate may be just beginning (see **[38]** and **[50]**).

For amplification of the issues examined in this section see **[1]**, **[8]**, **[11]**, **[12]**, **[14]**, **[15]**, **[29]**, **[30]**, **[38]**–**[43]**, **[47]**–**[51]**, **[54]**, **[55]**, **[58]**, **[62]**–**[64]**.

REFERENCES

1. K. Appel and W. Haken, The four-color problem, *Mathematics Today—Twelve Informal Essays*, ed. by L. A. Steen, Springer-Verlag New York, Inc., 1978, pp. 153–190.
2. E. Bishop, *Foundations of Constructive Analysis*, McGraw-Hill Book Co., New York, 1967.
3. U. Bottazzini, *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*, Springer-Verlag New York, Inc., 1986.
4.. N. Bourbaki, The architecture of mathematics, *Amer. Math. Monthly* 57 (1950), 221–232.
5. D. Bridges and F. Richman, *Varieties of Constructive Mathematics*, Cambridge University Press, New York, 1987.
6. D. M. Burton, *The History of Mathematics*, Allyn and Bacon, Boston, 1985.
7. A. Caldar, Constructive mathematics, *Scientific Amer.* 241 (October 1979), 146–171.
8. E. J. Chaitin, Randomness and mathematical proof, *Scientific Amer.* 232 (May 1975), 47–52.
9. C. T. Chong and Y. K. Leong, An interview with Jean-Pierre Serre, *Math. Intell.* 8:4 (1986), 8–13.
10. B. A. Cipra, Computer search solves an old math problem, *Science* 242 (December 16, 1988), 1507–1508.
11. P. J. Davis, Fidelity in mathematical discourse: Is one and one really two?, *Amer. Math. Monthly* 79 (1972), 252–262.
12. P. J. Davis and R. Hersh, *The Mathematical Experience*, Birkhäuser, Boston, 1981.
13. R. Dedekind, *Essays on the Theory of Numbers*, Dover Publications, Mineola, NY, 1963 (orig. 1901).
14. R. A. De Millo, R. J. Lipton, and A. J. Perlis, Social processes and proofs of theorems and programs, *Communications of the ACM* 22 (1979), 271–280.
15. M. Detlefsen and M. Luker, The four-color theorem and mathematical proof, *J. of Phil.* 77 (1980), 803–820.

[52] Such results apparently can be applied with impunity to cryptography, which is the main field of application of primality testing. It is noteworthy, moreover, that the proofs of such results use highly sophisticated abstract mathematics such as abelian varieties and Faltings' results dealing with the Mordell conjecture. See **[39]** (which also contains an update of Rabin's work).

[53] It has subsequently been shown that this number is indeed a prime (**[50**, p. 102**]**).

16. C. H. Edwards, *The Historical Development of the Calculus*, Springer-Verlag New York, Inc., 1979.
17. S. Engelsman, *Families of Curves and the Origins of Partial Differentiation*, North-Holland, New York, 1984.
18. H. Eves, *Great Moments in Mathematics*, 2 vols. (before 1650, and after 1650, resp.), MAA, Washington, 1983.
19. S. Feferman, What does logic have to tell us about mathematical proofs?, *Math. Intell.* 2:1 (1979), 20–24.
20. C. G. Fraser, The calculus as algebraic analysis: some observations on mathematical analysis in the 18th century, *Arch. Hist. Ex. Sc.* 39 (1989), 317–335.
21. N. Goodman, Mathematics as an objective science, *Amer. Math. Monthly* 86 (1979), 540–551.
22. J. V. Grabiner, Is mathematical truth time-dependent? *Amer. Math. Monthly* 81 (1974), 354–365.
23. ______, Changing attitudes toward mathematical rigor: Lagrange and analysis in the 18th and 19th centuries, *Epistemological and Social Problems of the Sciences in the Early 19th Century*, ed. by H. Jahnke and M. Otte, D. Reidel, Boston, 1981, pp. 311–330.
24. ______, *The Origins of Cauchy's Rigorous Calculus*, M.I.T. Press, Cambridge, MA, 1981.
25. ______, Who gave you the epsilon: Cauchy and the origins of rigorous calculus, *Amer. Math. Monthly* 90 (1983), 185–194.
26. ______, The changing concept of change: The derivative from Fermat to Weierstrass, this MAGAZINE, 56 (1983), 195–206.
27. J. Gray, Review of *Über die Enstehung von David Hilberts "Grundlagen der Geometrie"*, *Hist. Math.* 15 (1988), 181–183.
28. I. Hacking, Proof and eternal truths: Descartes and Leibniz, *Descartes: Philosophy, Mathematics and Physics*, ed. by S. Gaukroger, Barnes and Noble Books, New York, 1980, pp. 169–180.
29. W. Haken, An attempt to understand the four-color problem, *J. Graph Theory* 1 (1977), 193–206.
30. G. Hanna, *Rigorous Proof in Mathematics Education*, Ontario Institute for Studies in Education Press, Toronto, 1983.
31. P. Kitcher, *The Nature of Mathematical Knowledge*, Oxford University Press, Inc, Fair Lawn, NJ, 1983.
32. I. Kleiner, Evolution of group theory: A brief survey, this MAGAZINE, 59 (1986), 195–215.
33. ______, A sketch of the evolution of (noncommutative) ring theory, *L'Enseign. Math.* 33 (1987), 227-267.
34. ______, Evolution of the function concept: A brief survey, *Coll. Math. J.* 20 (1989), 282–300.
35. M. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, Inc., Fair Lawn, NJ, 1972.
36. ______, *Mathematics: The Loss of Certainty*, Oxford University Press, Inc., Fair Lawn, NJ, 1980.
37. W. R. Knorr, On the early history of axiomatics: The interaction of mathematics and philosophy in Greek antiquity, *Theory Change, Ancient Axiomatics and Galileo's Methodology*, ed. by J. Hintikka et al., D. Reidel, Boston, 1980, pp. 145–186.
38. G. Kolata, Mathematical proofs: The genesis of reasonable doubt, *Science* 192 (June 1976), 989–990.
39. ______, Prime tests and keeping proofs secret, *Science* 233 (Aug. 1986), 938–939.
40. I. Lakatos, *Proofs and Refutations*, Cambridge University Press, New York, 1976.
41. ______, Cauchy and the continuum: The significance of non-standard analysis for the history and philosophy of mathematics, *Math. Intell.* 1:3 (1978), 151–161.
42. ______, A renaissance of empiricism in the recent philosophy of the mathematics? *New Directions in the Philosophy of Mathematics*, ed. by T. Tymoczko, Birkhäuser, Boston, 1986, pp. 29–48.
43. ______, What does a mathematical proof prove?, *New Directions in the Philosophy of Mathematics*, ed. by T. Tymoczko, Birkhäuser, Boston, 1986, pp. 153–162.
44. D. Laugwitz, Infinitely small quantities in Cauchy's textbooks, *Hist. Math.* 14 (1987), 258–274.
45. G. E. R. Lloyd, *Magic, Reason and Experience: Studies in the Origin and Development of Greek Science*, Cambridge University Press, New York, 1979.
46. M. Mandelkern, Constructive mathematics, this MAGAZINE, 58 (1985), 272–280.
47. Yu. I. Manin, How convincing is a proof? *Math. Intell.* 2:1 (1979), 17–18.
48. A. R. Meyer, The inherent computational complexity of theories of ordered sets, *Proc. Int. Congr. of Mathematicians*, Vancouver, 1974, 477–482.
49. F. H. Norwood, Long proofs, Vol. 2 , *Amer. Math. Monthly* 89 (1982), 110–112.
50. C. Pomerance, Recent developments in primality testing, *Math. Intell.* 3:3 (1981), 97–105.
51. M. O. Rabin, Probabilistic algorithms, *Algorithms and Complexity: New Directions and Recent Results*, ed. by J. F. Traub, Academic Press, New York, 1976, pp. 21–40.
52. G. F. Simmons, *Differential Equations*, McGraw-Hill Book Co., New York, 1972.
53. S. Smale, Algebra and complexity theory, *Bull. Amer. Math. Soc.* 4 (1981), 1–36.
54. J. Spencer, Short theorems with long proofs, *Amer. Math. Monthly* 90 (1983), 365–366.

55. L. A. Steen, Living with a new mathematical species, *Math. Intell.* 8:2 (1986), 33–40.

56. L. J. Stockmeyer and A. K. Chandra, Intrinsically difficult problems, *Scientific Amer.* 240 (May 1979), 140–159.

57. D. J. Struik, *A Source Book in Mathematics, 1200–1800*, Harvard Univeristy Press, Cambridge, MA, 1969.

58. E. R. Swart, The philosophical implications of the four-color problem, *Amer. Math. Monthly* 87 (1980), 697–707.

59. A. Szabo, *The Beginnings of Greek Mathematics*, D. Reidel, Boston, 1978.

60. A. Tarski, Truth and proof, *Scientific Amer.* 220 (June 1969), 63–77.

61. *The Two-Year College Mathematics Journal* 12:2 (March 1981)—this issue features the concept of proof. It includes articles by Appel and Haken, Galda, Renz, and Tymoczko.

62. T. Tymoczko, The four-color problem and its philosophical significance, *J. of Phil.* 76:2 (1979), 57–83.

63. ______, Computers, proofs and mathematicians: a philosophical investigation of the four-color proof, this MAGAZINE, 53 (1980), 131–138.

64. ______, Making room for mathematicians in the philosophy of mathematics, *Math. Intell.* 8:3 (1986), 44–50.

65. B. L. Van der Waerden, *Science Awakening*, P. Noordhoff, Groningen, 1954.

66. J. Von Neumann, The mathematician, *The World of Mathematics*, Vol. 4, ed. by J. R. Newman, Simon and Schuster, New York, 1956, pp. 2053–2063.

67. ______, The role of mathematics in the sciences and in society, *Collected Works*, Vol. 6, ed. by A. H. Taub, Macmillan Publishing Co., New York, 1963, pp. 477–490.

68. H. Weyl, Mathematics and logic, *Amer. Math. Monthly* 53 (1946), 2–13.

69. ______, Axiomatic versus constructive procedures in mathematics, *Math. Intell.* 7:4 (1985), 10–17.

70. R. L. Wilder, The role of the axiomatic method, *Amer. Math. Monthly* 74 (1967), 115–127.

71. ______, *Evolution of Mathematical Concepts*, John Wiley and Sons, Inc., New York, 1968.