INDEX TO VOLUME 64

AUTHORS

TITLES

PROBLEMS

P, S, Q, *and* C *refers to Proposals, Solutions, Quickies, and Comments, respectively. Page numbers are given in parentheses. Thus,* P1338(55) *refers to proposal number* 1338, *which appears on page* 55.

February: P1363-1367; Q772-774; S1338-1342.
April: P1368-1372; Q775-777; S1343-1347.
June: P1373-1377; Q778-780; S1348-1352.
October: P1378-1382; Q781-783; S1353-1357.
December: P1383-1387; Q784-785; S1358-1362.

NEWS AND LETTERS

Vol. 64 No. 5, December 1991

MATHEMATICS MAGAZINE

- Rigor and Proof in Mathematics: A Historical Perspective
- Serial Isogons of 90 Degrees
- Which Rectangular Chessboards Have a Knight's Tour?

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

The aim of *Mathematics Magazine* is to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these are essential to the presentation of some good mathematics. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

The full statement of editorial policy appears in this *Magazine*, Vol. 64, pp. 71–72, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher.

Send new manuscripts to: Martha Siegel, Editor, Mathematics Magazine, Towson State University, Towson, MD 21204. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and two copies and keep one copy. In addition, authors should supply the full five-symbol Mathematics Subject Classification number, as described in *Mathematical Reviews*, 1980 and later. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added. Do not use staples.

Cover illustration: Carolyn Westbrook

AUTHOR

Israel Kleiner received his Ph.D. in ring theory at McGill University, and has been at York University for more than 20 years. He is coordinator of an in-service Master's Programme in mathematics for teachers, and is an active secondary-school liaison. The history of mathematics and its role in the teaching of mathematics are among his major interests.

Vol. 64 No. 5, December 1991

MATHEMATICS MAGAZINE

The *MATHEMATICS MAGAZINE* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $16 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $64. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $68 per year.

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, The Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

Second class postage paid at Washington, D.C. and additional mailing offices.

Postmaster: Send address changes to Mathematics Magazine Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036-1385.

PRINTED IN THE UNITED STATES OF AMERICA

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 $dx\,dy$ ... 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

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

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

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

The above equation becomes

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

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

$$a_n = \frac{1}{l}\int_{-l}^{l} f(t)\cos\frac{n\pi t}{l}\,dt{,}2s2s b_n = \frac{1}{l}\int_{-l}^{l} f(t)\sin\frac{n\pi t}{l}\,dt.$$

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

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

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.

NOTES

Serial Isogons of 90 Degrees

LEE SALLOWS
Buurmansweg 30
6525 RW Nijmegen, The Netherlands

MARTIN GARDNER
110 Glenbrook Drive
Hendersonville, NC 28739

RICHARD K. GUY
University of Calgary
Calgary, Alberta, Canada T2N 1N4

DONALD KNUTH
Stanford University
Stanford, CA 94305

In 1988 the first author devised a computer program to search a unit-square grid for closed paths with the following properties. The path starts along a lattice line with a segment of unit length, turns 90 degrees in either direction, continues for 2 units, turns again in either direction, continues for 3 units, and so on. In other words, the segments of the path are in serial order $1, 2, 3, \ldots N$, with a right angle turn at the end of each segment. A path of N segments—the number is of course the same as the number of turns or corners—is said to be a path of order N.

If the path returns to its starting point, making a right-angle with its first segment, we call it a *serial isogon* of 90 degrees. The isogon is allowed to self-intersect, to touch at corners, and to overlap along segments. Think of it as a serial walk through a city of square blocks and returning to the starting corner, or as the moves of a rook on a large enough chessboard.

It is not obvious that such paths exist. However, with a little doodling you will discover the unique isogon of order 8, the lowest order a 90-degree isogon can have. It outlines a polyomino of 52 unit squares that, as FIGURE 1 shows, tiles the plane. Indeed, it satisfies the "Conway criterion" [1] for identifying tiling shapes. The boundary of the polyomino can be partitioned into six parts, the first and fourth of which (AB and ED in FIGURE 2) are equal and parallel, while the other four ($BC = 3$; $CD = 4, 5, 4$; $EF = 7$; $FA = 1, 8, 1$) each have rotational symmetry through 180° about their midpoints (black dots in FIGURE 2). It may be the only plane-tiling polyomino with a serial boundary, but we are unable to prove this.

Puzzle: Can you tile this polyomino with thirteen L-shaped tetrominoes?

It is easy to prove that N must be a multiple of 4. One way is to consider N rook moves on a bicolored chessboard. Assume without loss of generality that the rook begins on a black cell and makes its first move horizontally. To close the path, its final move must be vertical, and end on a black square. Because moves of odd and even length alternate, the sequence of colors at the end of each segment forms the repeating sequence: $WW\ BB\ WW\ BB\ldots$. The rook returns to a black cell, after a vertical move, if and only if the number of moves $\equiv 0 \bmod 4$.

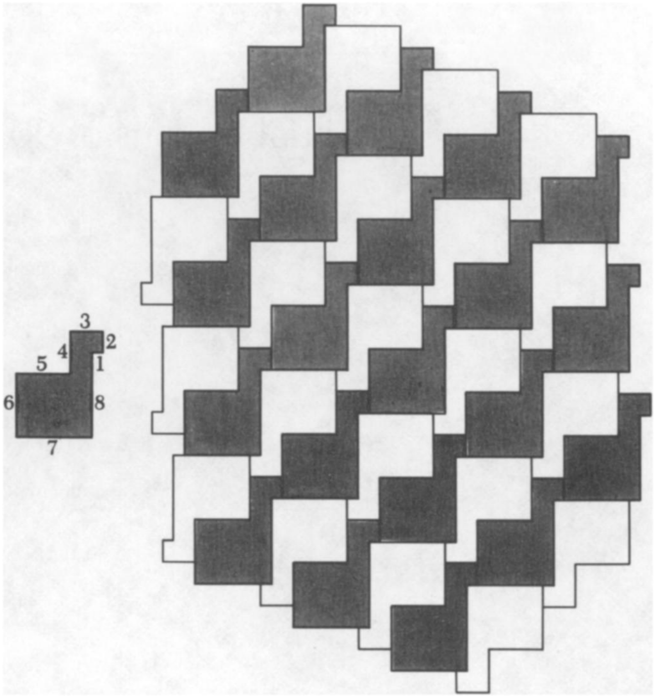


FIGURE 1
The only known serial-sided polyomino that tiles the plane.

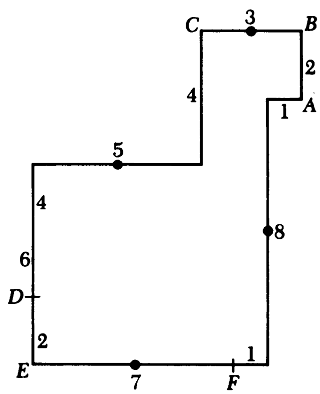


FIGURE 2
Applying Conway's criterion to prove that the polyomino tiles the plane.

Experimenting on graph paper will quickly convince you that there is no serial isogon of order 4. (You might call this "a 4-gon conclusion".) We have exhibited one of order 8. After the computer program exhaustively plotted all serial isogons of orders through 24, a surprising fact emerged. No serial isogons were found except when N is a multiple of 8. This led to several proofs of the following theorem:

For any 90-*degree serial isogon*, N *must be a multiple of* 8.

Assume that a closed path begins with a unit move to the east, and that moving east or north is positive, and moving south or west is negative. A path can be uniquely described by placing a plus or minus sign in front of each number in the sequence of moves to indicate the direction of the move. For example, the order-8 polyomino has the following formula:

$$+1+2-3-4-5-6+7+8.$$

It is obvious that if the path closes, the sum of all horizontal moves—the odd numbers—must be zero, otherwise the path will not return to the vertical lattice line that goes through the starting point. Similarly, the sum of all vertical moves—even numbers—must be zero or the path will not return to the horizontal lattice line going through the origin point. The sum of all the numbers will, of course, also be zero.

We know that N is a multiple of 4, say $4k$. Then the north-south moves are the even-length ones, $2, 4, \dots, 4k$. The total north-south distance is therefore $2(1+2+\cdots+2k) = 2k(2k+1)$. Half of this, $k(2k+1)$, must be north and half of it south. But if k is odd, this distance is odd, and cannot be the sum of even-length moves.

We can make this clearer by taking $N = 12$ as an example. Even numbers in this path's formula $(2, 4, 6, 8, 10, 12)$ add to 42. If the formula describes a closed path, the sum of the positive numbers in this sequence must equal $42/2 = 21$. But no set of even numbers can add to the odd number 21. Consequently, no formula can be constructed that will describe a closed path of order 12.

With reference to the grid, this tells us that if N is a multiple of 4 but not of 8, the last segment of the path, which is vertical, cannot return to the horizontal lattice line that goes through the path's origin point because the sum of the positive segments going north cannot equal the absolute sum of the negative segments going south. The path's end will always be an even number of units above or below the zero horizontal line.

Is there a coloring pattern that proves the $N \equiv 0 \bmod 8$ theorem? Yes, the simple coloring shown in FIGURE 3 (found by the second author) will do the trick.

Start a path on any black cell, then make your first move horizontally in either direction. Regardless of your choices of how to turn at the end of each segment, the colors at the ends of segments will endlessly repeat the sequence: *WWWWWWBB*,

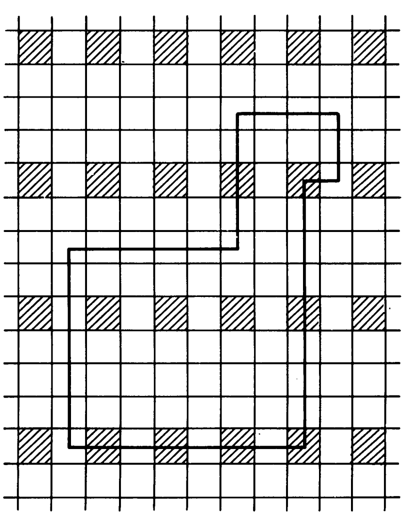


FIGURE 3
A bicoloring of the grid for proving that N is a multiple of 8 for all orders of closed serial paths. The order-8 polyomino is shown in outline.

WWWWWWWBB, *WWWWWWWBB*, ... The path will enter a black cell (which it must do if the path closes) in a direction perpendicular to the first segment if and only if the number of moves is a multiple of 8. Shown in the illustration is the path that outlines the order-8 tiling polyomino.

Although all closed paths have formulas in which the total sum of the signed terms is zero, this is not sufficient to produce a formula for a path. A serial path closes if and only if the even numbers in its formula add to zero, and likewise the odd numbers. (Nonserial closed paths may meet this proviso, and be of any order of 4 or greater that is a multiple of 2. The formula $+1+2-1-2$, for example, describes a closed path that outlines a domino.) If this proviso is not met, the formula gives the location of the path's final point with respect to the origin. If the sum of the signed even numbers is positive, it gives the number of units where the path ends above the origin; if negative, it gives the number of units below the origin. Similarly for the sum of the signed odd numbers. If positive, it gives the end point's distance east of the origin; if negative, the distance west.

If the signs of all even numbers are changed, or if the signs of all odd numbers are changed, it reflects the isogon along an orthogonal axis. If all signs are changed, it rotates the isogon 180 degrees.

We have shown that $N \equiv 0 \bmod 8$ is a necessary condition for a closed serial path. Is it also sufficient? Yes. Here is one way to arrange plus and minus signs in a formula that will always describe a serial isogon: Put plus signs in front of the first and last pairs of numbers. Put minus signs in front of the next to last pairs of numbers at each end, and continue in this way until all pairs of numbers are signed. This ensures that all even numbers add to zero, and likewise all odd numbers, therefore the formula must describe a closed path. It produces, for instance, the unique formula for $N = 8$. Applied to $N = 16$ it gives the formula $+1+2-3-4+5+6-7-8-9-10+11+12-13-14+15+16$, which describes the isogon at position O_2E_5 in FIGURE 4.

Is there a procedure guaranteed to construct a serial isogon for any order $N \equiv 0 \bmod 8$ that outlines a polyomino? The answer is again yes. Each formula has $8n$ numbers. If we make positive all numbers in the first fourth of the formula, and also in the last fourth, and make negative all numbers in the half in between, we produce a serial isogon. Applied to $N = 16$, it gives $+1+2+3+4-5-6-7-8-9-10-11-12+13+14+15+16$, which describes the polyomino at O_1E_4 in FIGURE 4. The polyomino generated by this procedure always takes the form of a snake that grows longer as N increases. FIGURE 5 shows the polyomino snake for order 32. A diagonal line, running from the extreme corners of the snake's head and tail, going between all the interior corners, is almost, but not quite, straight.

We now turn to the more difficult task of enumerating all possible serial isogons (not counting rotations and reflections as different) for a given isogon. As mentioned earlier, there is only one isogon of order 8, the tiling polyomino. For $N = 16$, the computer program found the 28 solutions shown in FIGURE 4. Note that only three (O_1E_1, O_1E_4, O_1E_7) are polyominoes. For $N = 24$, the program produced 2,108 distinct isogons, of which 67 bound polyominoes. For $N = 32$ the program's running time became too long to be feasible.

No formula is known for enumerating all distinct serial isogons of order N, or for counting the polyominoes of a given order. However, there are procedures by which the number of isogons can be counted by hand to a value of N that goes well beyond $N = 24$.

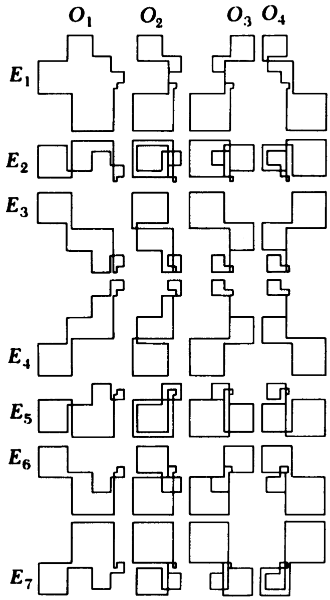


FIGURE 4
The 28 distinct order-16 serial isogons of 90°.

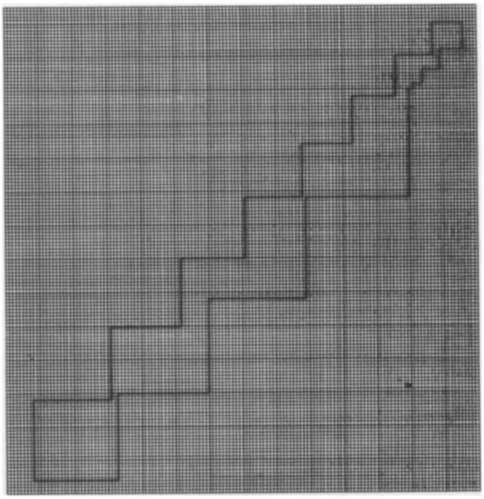


FIGURE 5
The snake polyomino of order 32.
The diagonal line separating the snake's two sides is almost, but not quite, straight.

Here is how the fourth author describes one such procedure:

Suppose $N=16$; it's easy to see that this quantity is the constant term if you expand the algebraic product

$$(x^{-1}+x^{1})(y^{-2}+y^{2})(x^{-3}+x^{3})\cdots(x^{-15}+x^{15})(y^{-16}+y^{16})$$

into powers of x and y. To get the number of ways for the odd sum to cancel, we want the constant term of

$$(x^{-1}+x^{1})(x^{-3}+x^{3})\cdots(x^{-15}+x^{15}),$$

which is the coefficient of $x^{1+3+5+\cdots+15}=x^{64}$ in

$$(1+x^{2})(1+x^{6})\cdots(1+x^{30}),$$

which is the coefficient of x^{32} in

$$(1+x)(1+x^{3})\cdots(1+x^{15}),$$

which is 8. To get the number of ways for the even sum to cancel, we want the constant term of

$$(y^{-2}+y^{2})(y^{-4}+y^{4})\cdots(y^{-16}+y^{16}),$$

which is the constant term of

$$(y^{-1}+y^{1})(y^{-2}+y^{2})\cdots(y^{-8}+y^{8}),$$

which is the coefficient of $y^{1+2+\cdots+8}=y^{36}$ in

$$(1+y^{2})(1+y^{4})\cdots(1+y^{16}),$$

which is the coefficient of y^{18} in

$$(1+y)(1+y^2)\cdots(1+y^8),$$

which is 14. So the total number of closed paths is 8×14; divide by 4 to get 28 closed paths that are distinct under reflectional symmetry. Suppose we start with $+1+2$; then the four ways to do the odd numbers are

$$\begin{aligned} O_1 &= +1+3-5-7-9-11+13+15 \\ O_2 &= +1-3+5-7-9+11-13+15 \\ O_3 &= +1-3-5+7+9-11-13+15 \\ O_4 &= +1-3-5+7-9+11+13-15 \end{aligned}$$

and the seven ways to do the even numbers are

$$\begin{aligned} E_1 &= +2+4+6+8-10-12-14+16 \\ E_2 &= +2+4+6-8+10-12+14-16 \\ E_3 &= +2+4-6+8+10+12-14-16 \\ E_4 &= +2+4-6-8-10-12+14+16 \\ E_5 &= +2-4+6-8-10+12-14+16 \\ E_6 &= +2-4-6+8+10-12-14+16 \\ E_7 &= +2-4-6+8-10+12+14-16. \end{aligned}$$

The three serial polyominoes are the snake O_1E_4 and two other solutions O_1E_1, O_1E_7. (It's curious that only O_1 can be completed. The case O_2E_1 almost works, but that path gives a degenerate polyomino whose width is zero at one point. Paths O_3E_1, O_3E_4, and O_4E_4 fail in the same way.)

In general when $N = 8n$, the number of closed paths is the product of the coefficient of x^{8n^2} in

$$(1+x)(1+x^3)(1+x^5)\cdots(1+x^{8n-3})(1+x^{8n-1})$$

and the coefficient of y^{4n^2+2} in

$$(1+y)(1+y^2)(1+y^3)\cdots(1+y^{4n-1})(1+y^{4n}).$$

These numbers, for small n (divided by 2 to remove symmetry), are

N	n	odds/2	evens/2	product
8	1	1	1	1
16	2	4	7	28
24	3	34	62	2108
32	4	346	657	227322
40	5	3965	7636	30276740
48	6	48396	93846	4541771016
56	7	615966	1199892	739092675672
64	8	8082457	15796439	127674038970623

It seems certain that the vast majority of these isogons will not bound polyominoes. The paper of Bhattacharya and Rosenfeld [3] is concerned with the problem of avoiding self-intersections in isogons: they treat the general problem in which the sides are of arbitrary length, not just our particular case of consective integers.

Here is how the third author has made an asymptotic estimate of the number of serial isogons:

We have seen that the number of isogons of a given order is the product of half the number of possible choices of sign in $\pm 2 \pm 4 \pm 6 \pm \cdots \pm (8n-2) \pm 8n = 0$ with half the number of choices of sign in $\pm 1 \pm 3 \pm 5 \pm \cdots \pm (8n-3) \pm 8n = 0$.

The first of these is the number of partitions of half of the sum

$$2 + 4 + 6 + \cdots + 8n$$

into distinct even parts, of size at most $8n$, i.e., the number of partitions of $n(4n+1)$ into d distinct parts of size $\leq 4n$. Subtract $1, 2, \ldots, d$ from these parts, now no longer necessarily distinct, of size $\leq 4n - d$. We require the number of partitions of $4n^2 + n - (1/2)d(d+1)$ into at most d parts, no longer necessarily distinct, of size $\leq 4n - d$. Here d, the number of parts, lies in the approximate range $(4 - 2\sqrt{2})n < d < 2\sqrt{2}\,n$.

In the same way the second number is the number of partitions of $4n^2 - (1/2)d^2$ into at most d parts, not necessarily distinct, of size $\leq 4n - d$, with d in approximately the same range as before, but with d necessarily even!

The main contribution comes from $d = 2n$ and the distribution, as we shall see, is essentially the binomial distribution, so that a good estimate of the whole is obtained by multiplying this central term by $\sqrt{4\pi n}$, except that we halve the "odd" estimate since only alternate terms (d even) are taken.

From formula (75) in [2] we learn that the number of partitions of j into at most a parts, with each part $\leq b$, is asymptotically equal to

$$\frac{1}{\sigma_{a,b}}\binom{a+b}{a}\phi\left(\frac{j - ab/2}{\sigma_{a,b}}\right),$$

where $\sigma_{a,b}^2 = ab(a+b+1)/12$ and $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$.

In both the even and the odd cases, $a = d$, $b = 4n - d$, $a + b + 1 = 4n + 1$, and the central term is given by $d = 2n = a = b$ for which $j = 4n^2 + n - n(2n+1)$ and $j = 4n^2 - 2n^2$, i.e., $2n^2$ in either case, and $j - ab/2 = 0$, so that the central term is asymptotically equal to

$$\frac{1}{\sqrt{2\pi}}\,\frac{\sqrt{3}}{2n\sqrt{4n+1}}\binom{4n}{2n},$$

i.e., asymptotically equal to

$$\frac{\sqrt{3}}{2n\sqrt{2\pi}\,\sqrt{4n+1}}\,\frac{2^{4n}}{\sqrt{4\pi n}}$$

by Stirling's formula. Multiply by $\sqrt{4\pi n}$ to estimate the total number of partitions in the even case, and by half of that in the odd case. The product of the halves of these numbers (i.e., not counting the E.-W. or N.-S. reflections of the isogons as different) is thus

$$\frac{1}{8}\left\{\frac{\sqrt{3}}{2n\sqrt{2\pi}\,\sqrt{4n+1}}2^{4n}\right\}^2 = \frac{3 \cdot 2^{8n-6}}{\pi n^2(4n+1)}.$$

Compare this estimate with the actual values obtained above.

n	$3 \times 2^{8n-6}/\pi n^2(4n+1)$
1	0.76
2	27.2
3	2140
4	235604
5	31248698
6	4666472281
7	756618728785
8	130321844073100

The first author has since explored serial isogons on isometric grids. They are of two types: those with 60-degree angles, and those with 120-degree angles. In the 60-degree case, serial isogons exist if and only if $N \geq 9$ and $N \not\equiv 1 \bmod 3$. In the 120-degree case, they exist if and only if N is a multiple of 6. If each angle of a serial polygon can be either 60 or 120 degrees, such polygons exist for any N that is 5 or greater. In all three cases, the smallest order is unique and is a *polyiamond*—a polygon formed by joining unit equilateral triangles. The three are shown in FIGURE 6. Note that the order-5 polyiamond tiles the plane in two different ways, see [4].

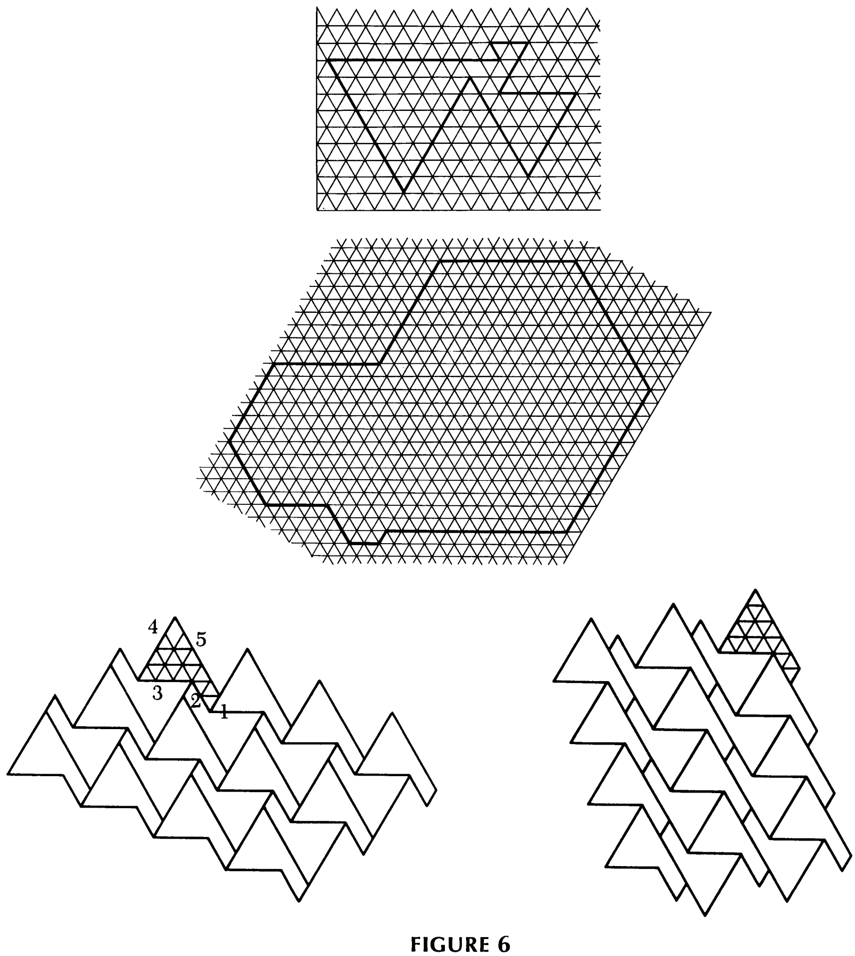


FIGURE 6
The unique, smallest examples of serial isometric isogons. At top is the only order-9 isogon with 60-degree angles. In the middle is the only order-12 isogon with 120-degree angles. At bottom, shown tiling the plane in two ways, is the only order-5 isogon that mixes the two angles.

For a time the first author believed that closed serial paths existed only when angles are 60, 90, or 120 degrees, but then he suddenly discovered such a path with 108-degree angles, the angles of a regular pentagon. More recently Hans Cornet, a retired mathematician in The Hague, has put forward a proof that at least one serial isogon can be constructed for every angle α that is a rational multiple of 360 degrees, that is, when $\alpha = 2\pi n/m$ radians, with m and n positive integers; see [4].

The first author has also investigated closed paths on square and isometric grids that have segment lengths in sequences other than the counting numbers, such as Fibonacci sequences, consecutive primes, and so on. He has produced a whimsical class of *piominoes*—polyominoes whose sides in cyclic order are the first n digits of pi, with zeros omitted. Because the digits of pi are pseudo-random, the task of enumerating pi-isogons of 90 degrees is related to determining the probability of self-intersecting random walks on a square lattice. He has even experimented with closed paths based on the letters of number words. An example is shown in FIGURE 7. It is unquestionably one of the most useless polyomino outlines ever constructed, yet does it not have a curious charm?

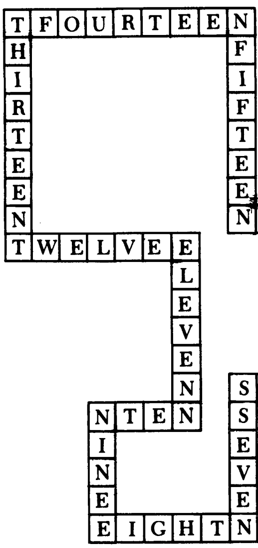


FIGURE 7
A 16-sided polyomino, its side determined by the words for zero through 15.

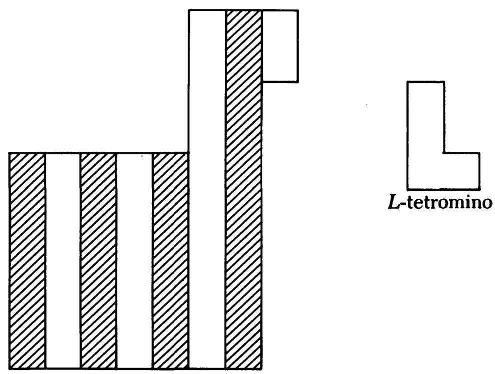


FIGURE 8
A coloring pattern on the tiling polyomino for proving it cannot be dissected into 13 *L*-tetrominoes.

Solution to puzzle: To prove that 13 *L*-tetrominoes will not tile the polyomino, divide the polyomino into alternatively colored vertical stripes as shown in FIGURE 8. No matter how an *L*-tetromino is placed within this pattern, it will cover an odd number of cells of each color. Thirteen odd numbers add to an odd number, but the polyomino has an even number of cells of each color. The tiling is therefore impossible.

Postscript

The appearance of a prepublication version of this paper [4] has resulted in a good deal of correspondence and misunderstanding. John Leech observes that the argument there, showing that N is a multiple of 8, is faulty and exacerbated by the statement that "even numbers hardly entered into the discussion," whereas we know that it is the even lengths that clinch the matter. Leech also gives an argument involving those gridlines that "quarter" a tetromino, which is equivalent to that of FIGURE 8. The difficulty of printing a complicated formula in a popular article resulted in an oversimplification that was no longer an asymptotic formula in the sense described. Finally, it was implied that calculation of the exact numbers of serial isogons was more difficult than is actually the case, which prompted several people to reach for their computers. Ilan Vardi [6] has explained how to compute the numbers rapidly via the Chinese Remainder Theorem, and he lists the values for $N = 400$ and $N = 1000$. Calculations up to $N = 200$ were carried out by Sivy Farhi, 815 S. California #B, Monrovia, CA and by Pierre Barnouin, Chemin de Stramousse, 06530 Cabris, France, who gave the following values:

16	28
24	2108
32	227322
40	30276740
48	4541771016
56	739092675672
64	127674038970623
72	23085759901610016
80	4327973308197103600
88	835531767841066680300
96	165266721954751746697155
104	33364181616540879268092840
112	6854017416098227836106023048
120	1429368258586343246184813682344
128	302023498629081603279538134332922
136	64557914743374337032608546756101824
144	13941247125893997584457711273087122310
152	3038225349257507092516361163813831321438
160	667575475791956832191676953455074834982100
168	147773788473936923724715382248726990582150405
176	329316592421079646570222645485385251429569140 56
184	7383987729780296585063944629621065123001927478725
192	166496155571027370972412626231396934148397663305 8164
200	3773597098720565621984238570532885702325776079874434 92

REFERENCES

1. Doris Schattschneider, Will it tile? try the Conway criterion!, this MAGAZINE 53 (1980), 224–232.
2. Lajos Takács, Some asymptotic formulas for lattice paths, *J. Statist. Planning and Inference* 14 (1986), 123–142.
3. Prabir Bhattacharya and Azriel Rosenfeld, Contour codes of isothetic polygons, *Computer Vision, Graphics, and Image Processing* 50 (1990), 353–363.
4. Lee Sallows, New pathways in serial isogons, *The Mathematical Intelligencer* (in press).
5. A. K. Dewdney, Mathematical Recreations: An odd journey along even roads leads to home in Golygon City, *Sci. Amer.* 263 #1 (July 1990), 118–121.
6. Ilan Vardi, *Computational Recreations in Mathematica*, Addison-Wesley, Reading, MA, 1991, Section 5.3.

Which Rectangular Chessboards Have a Knight's Tour?

ALLEN J. SCHWENK*
Western Michigan University
Kalamazoo, MI 49008

Problems involving the search for Hamiltonian cycles are popular in undergraduate discrete mathematics courses. A few textbooks introduce the intriguing puzzle of searching for spanning tours by a knight on various rectangular chessboards. This area provides a down-to-earth collection of problems that illustrates the idea of a Hamiltonian cycle. The problems are challenging enough to require thoughtful solutions, and yet, at least for small boards, manageable enough so that students can succeed in finding tours on some boards and in showing that they are impossible on others. It also gives the instructor a chance to prove the nonexistence of tours on an infinite family of boards by an elegant (though well-known) parity argument. Certainly any curious student must wonder precisely which size boards do admit knight's tours and which do not. Chartrand [2] ignores this natural question, while Wilson and Watkins [7] report that the question was fully resolved by Euler in 1759 and 12 years later (independently) by Vandermonde. Similarly, Berge [1] introduces the problem, mentions some of the history, and then immediately drops it. Dudeney [3] also provides a sketchy history. Rouse Ball and Coxeter [6] provide a 10-page treatment of the problem without ever mentioning which size boards can in fact be toured. A recent research article by Eggleton and Eid [4] focuses on "open" tours for which the knight need not return to his starting square. They even extend the problem to infinite boards of various types, leading to intriguing questions about the existence of spanning one-way and two-way infinite paths. But their discussion of the original knight's tour problem only goes into detail on the well-known odd order case and on the family of $3 \times n$ boards where they report a private communication claiming (erroneously) that Hamiltonian cycles exist if and only if $n \geqslant 8$ and n is even. We shall show that the correct version is $n \geqslant 10$ and n is even. The universal avoidance of reporting the definitive solution creates the impression that it must be beyond the undergraduate level. Presumably, it is difficult to describe the sizes that admit a tour, harder still to actually construct these tours, and heaven knows what it takes to show that all other sizes really are impossible. The 200-year-old references to the literature are incomplete and intimidating. I don't know how to find these ancient volumes. My students wouldn't even consider trying.

The purpose of this article is to show that the full solution of the knight's tour problem is quite brief and entirely accessible to beginning students. In the process, the student will see a new use of parity to show impossibility in one case and a rather unusual instance of proof by induction that requires nine specific cases in order to anchor the induction.

We begin with a careful definition of the problem. An $m \times n$ chessboard is an array with square cells arranged in m rows and n columns. The standard chessboard is 8×8. For convenience we shall assume $m \leqslant n$. We label the cells (i, j) counting from the upper left corner in matrix fashion. Now a legal knight move is the result of moving two cells horizontally or vertically and then turning and moving one cell in

*Research supported in part by the Office of Naval Research, Contract N0014-91-J-1364.

the perpendicular direction. Thus, if we start at cell (i, j) we can complete the move on one of eight cells: $(i \pm 2, j \pm 1)$ or $(i \pm 1, j \pm 2)$. Of course if we are too close to the border of the board some of these choices may not exist. The knight's tour question is usually posed in this form:

Problem. On which $m \times n$ boards can a knight make successive legal knight moves, visit every cell exactly once, and conclude by returning to its starting cell?

There is also a version of the problem seeking "open tours" where the knight is not required to return to his starting position. The open tour problem can be solved by the same methods as the more common "closed tour" problem; we shall leave it as a challenge for the interested reader. The first step is to convert the problem to a question about certain graphs. We define a graph $G(m, n)$ on mn vertices by replacing each cell of the board by a vertex and then joining two vertices by an edge if they are separated by a knight's move. This is illustrated for a 3×6 board in FIGURE 1. A knight can tour the $m \times n$ board if and only if there exists a cycle containing all the vertices in the resulting graph. Such a cycle is called a Hamiltonian cycle, named after William R. Hamilton who marketed a puzzle called *A Voyage Round the World* based on this concept in 1859. Accounts of Hamilton's puzzle can be found in [2, 6, 7]. The customary alternating white and black coloring of the chessboard is preserved in the white and black vertices of the graph. We set vertex (i,j) to be white if $i + j$ is even and black if $i + j$ is odd. It is easy to see that every edge in the graph joins vertices of opposite colors. Such a graph is called *bipartite*, or for brevity, a *bigraph*. Since the colors must alternate in any cycle, the cycle must have an even number of vertices. We have just proved one of the first theorems on bipartite graphs, namely, all cycles must be even.

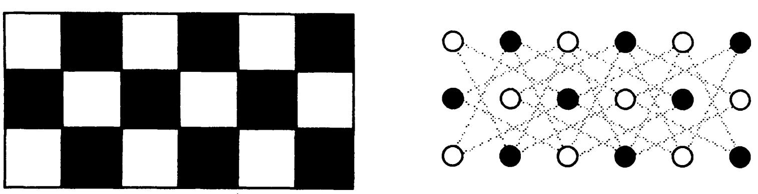


FIGURE 1
Conversion of the 3×6 chessboard into the graph $G(3, 6)$.

We can now state the conditions that determine which chessboards have a knight's tour.

THEOREM. *An $m \times n$ chessboard with $m \leqslant n$ has a knight's tour unless one or more of these three conditions holds:*
(a) *m and n are both odd;*
(b) *$m = 1$, 2, or 4; or*
(c) *$m = 3$ and $n = 4$, 6, or 8.*

Proof. We begin by showing why conditions (a), (b), and (c) must be excluded. Then we shall show how to construct a tour on every other board.

When m and n are both odd, so is the order, mn, of our graph. But we have already observed that every cycle must be even, and so no Hamiltonian cycle can exist.

For $m = 1$ or 2, it is clear that the board is not wide enough to permit a tour. Indeed, cell (1, 1) doesn't even have two available edges to be used in the cycle. For $m = 4$, the impossibility is more subtle. We present here the proof discovered by

Louis Pósa as a teenager and reported in the classic book of Ross Honsberger [5]. Assume that we have found a Hamiltonian cycle $v_1 v_2 \dots v_{4n} v_1$. Let us recolor the vertices red and blue, with every vertex in rows 1 and 4 red and every vertex in rows 2 and 3 blue. This coloring no longer serves as the bipartition for the graph since some blue vertices are adjacent to other blue vertices, for example, (2, 1) and (3, 3). However, every red vertex is adjacent only to blue vertices. Thus, in a presumed Hamiltonian cycle, the red vertices must always be separated by blue vertices. Since we have $2n$ vertices of each color, the red and blue vertices must alternate around the cycle. Now starting at $v_1 = (1, 1)$, we can conclude that all the vertices in odd positions on the cycle, v_{2k+1}, are red. But from the original black and white coloring we can conclude equally well that all the vertices v_{2k+1} are also white. Thus all red vertices are white vertices, but this contradicts the different pattern chosen for the two colorings. We conclude that no Hamiltonian cycle is possible.

To analyze condition (c) we introduce certain graphical concepts. The 3×4 board has already been excluded in the preceding paragraph. When we remove a vertex v from a graph G we also remove all edges incident with v. For any G having a Hamiltonian cycle, it is clear that removing any set of k vertices can leave at most k connected components. Since removing vertices (1, 3) and (3, 3) from $G(3,6)$ leaves three components, we must conclude that no Hamiltonian cycle exists for the 3×6 board. Now it happens that in $G(3,8)$ vertices (1, 1), (2, 1), (3, 1), (2, 2), (1, 8), (2, 8), (3, 8), and (2, 7) all have degree two, forcing us to include in a presumed Hamiltonian cycle the 16 edges shown in FIGURE 2. These edges form six paths that must lie within the Hamiltonian cycle. We also consider the two vertices missed by all six paths, namely (2, 4) and (2, 5) as trivial paths. We define a new graph $G^*(3,8)$ derived from these eight paths by letting one new vertex stand for each of these eight paths and joining two of these new vertices i and j whenever there is an edge in $G(3,8)$ joining an end of path i to an end of path j. Now a Hamiltonian cycle in the original $G(3,8)$ must force a corresponding Hamiltonian cycle to be present in $G^*(3,8)$, although the converse need not be true. But $G^*(3,8)$ has two vertices of degree three whose removal leaves three components. Therefore, neither $G^*(3,8)$ nor $G(3,8)$ can have a Hamiltonian cycle.

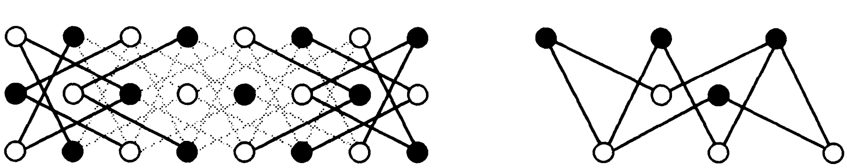


FIGURE 2

Sixteen edges that must belong to any Hamiltonian cycle of $G(3,8)$ and the resulting derived graph $G^*(3,8)$.

This completes the list of excluded sizes. Every other board has a Hamiltonian cycle, but how can we hope to construct all the necessary tours? The key is to develop a method that allows us to build new tours from smaller tours. In the following lemma it is convenient to dispense with the convention that $m \leqslant n$. This lemma allows us to add 4 columns to a successful tour, provided 10 particular edges belong to the tour. Actually, our extension methods may require one, two, or four particular edges to be present. The union of these sets gives five prescribed edges. But in order to be free to extend by either four columns or four rows, we require the presence of 10 specified edges.

LEMMA. *If $G(m,n)$ has a Hamiltonian cycle that includes the* 10 *edges*

$$(1,n-1)\text{–}(3,n) \quad (m-2,n-1)\text{–}(m,n) \quad (m-1,1)\text{–}(m,3) \quad (m-1,n-2)\text{–}(m,n)$$
$$(4,n-1)\text{–}(2,n) \quad (1,n)\text{–}(3,n-1) \quad (m-2,n)\text{–}(m,n-1) \quad (m,1)\text{–}(m-1,3)$$
$$(m,n-2)\text{–}(m-1,n) \quad (m,2)\text{–}(m-1,4),$$

then $G(m,n+4)$ also has a Hamiltonian cycle including the corresponding 10 *edges*

$$(1,n+3)\text{–}(3,n+4) \quad (m-2,n+3)\text{–}(m,n+4) \quad (m-1,1)\text{–}(m,3)$$
$$(m-1,n+2)\text{–}(m,n+4) \quad (4,n+3)\text{–}(2,n+4) \quad (1,n+4)\text{–}(3,n+3)$$
$$(m-2,n+4)\text{–}(m,n+3) \quad (m,1)\text{–}(m-1,3) \quad (m,n+2)\text{–}(m-1,n+4)$$
$$(m,2)\text{–}(m-1,4).$$

Proof. The 10 required edges are displayed visually in FIGURE 3. For $m = 3$, these "10 edges" degenerate into a set of seven. For all values of m and n, four of the 10 required edges (specifically the edges $(1,n)$–$(3,n-1)$, $(m,1)$–$(m-1,3)$, $(m-2,n-1)$–(m,n), and $(m-1,n-2)$–(m,n) that all lead into corner cells) are already forced to obtain any Hamiltonian cycle. Thus, the additional hypothesis needed to facilitate the induction is not as restrictive as it may at first appear. To add four columns to any Hamiltonian cycle in $G(3,n)$ that contains the critical seven edges, we place a certain 3×4 array with a spanning path along side $G(3,n)$, delete edge $(1,n-1)$–$(3,n)$ from the cycle, and insert edges $(1,n-1)$–$(2,n+1)$ and $(3,n)$–$(1,n+1)$ in order to incorporate the path into the cycle. FIGURE 4 shows the extension of a Hamiltonian cycle in $G(3,10)$ to one in $G(3,14)$ to illustrate this construction. The new Hamiltonian cycle also contains the prescribed seven edges, so it too can be used for further extensions.

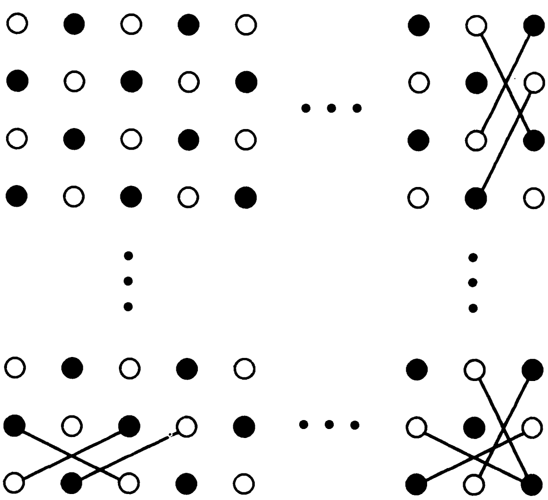


FIGURE 3
The ten edges required for the proposed induction.

For $m \geqslant 5$, we use an $m \times 4$ array $H(m,4)$ that is obtained from $G(m,4)$ by deleting all edges joining column two to column three and all edges joining vertices two columns apart except those joining vertices in rows 1 and 2 and those joining vertices in rows $m-1$ and m. The remaining graph $H(m,4)$ is regular of degree 2, that is, every vertex has degree 2. Its edges form cycles that wrap around the board hugging the outside border as closely as possible, but it is not a single cycle. It is easy to see that $H(m,4)$ has a pair of $2m$-cycles when m is odd and four m-cycles when m

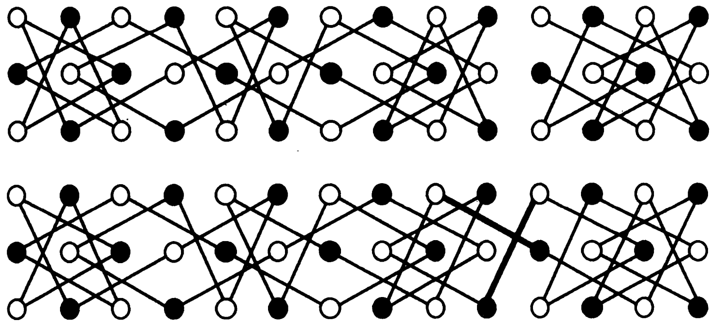


FIGURE 4
Extension of a Hamiltonian cycle in $G(3,n)$ to one in $G(3,n+4)$ for $n=10$.

is even. Since this fact is critical for the construction we are developing, we shall prove it by induction.

FIGURE 5 displays one of the cycles in $H(5,4)$. Its mate is formed by the reflection about the vertical axis through the center of the board. The additional two rows of vertices at the bottom suggest how the cycle shown is extended in $H(7,4)$. Specifically, we delete edges $(5,1)$–$(4,3)$ and $(5,2)$–$(4,4)$ and insert edges $(5,1)$–$(7,2)$, $(7,2)$–$(6,4)$, $(6,4)$–$(4,3)$, $(5,2)$–$(7,1)$, $(7,1)$–$(6,3)$, and $(6,3)$–$(4,4)$. Repeating this extension displays the paired-cycle structure in $H(m,4)$ whenever m is odd.

Similarly, FIGURE 6 displays two of the four cycles in $H(6,4)$. The other two mates are found by using a vertical reflection. Analogous to the odd case, we extend the cycles shown by deleting edges $(6,1)$–$(5,3)$ and $(6,2)$–$(5,4)$ and inserting edges $(6,1)$–$(8,2)$, $(8,2)$–$(7,4)$, $(7,4)$–$(5,3)$, $(6,2)$–$(8,1)$, $(8,1)$–$(7,3)$, and $(7,3)$–$(5,4)$. Repeating this extension displays the four cycle structure in $H(m,4)$ whenever m is even.

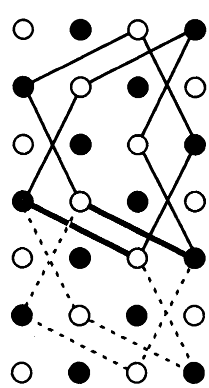


FIGURE 5
One of the cycles in $H(5,4)$ and its extension to $H(7,4)$.

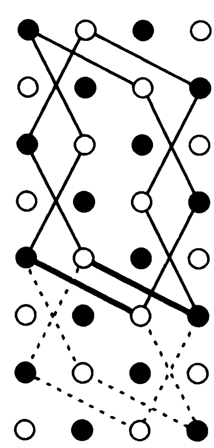


FIGURE 6
Two of the cycles in $H(6,4)$ and its extension to $H(8,4)$.

To extend a Hamiltonian cycle in $G(m,n)$ with m odd to one $G(m,n+4)$, we place $H(m,4)$ along side of $G(m,n)$ as shown in FIGURE 7. We remove two edges $(1,n)$–$(3,n-1)$ and $(2,n)$–$(4,n-1)$ from the Hamiltonian cycle, and two edges $(1,n+2)$–$(3,n+1)$ and $(2,n+2)$–$(4,n+1)$ from $H(m,4)$, and then insert four edges $(1,n)$–$(2,n+2)$, $(2,n)$–$(1,n+2)$, $(3,n-1)$–$(4,n+1)$, and $(4,n-1)$–$(3,n+1)$. This has the effect of incorporating the two cycles of $H(m,4)$ into the given Hamiltonian cycle to create a new Hamiltonian cycle in $G(m,n+4)$. The new cycle contains the prescribed 10 edges. The extension of $G(5,6)$ to $G(5,10)$ is illustrated in FIGURE 7.

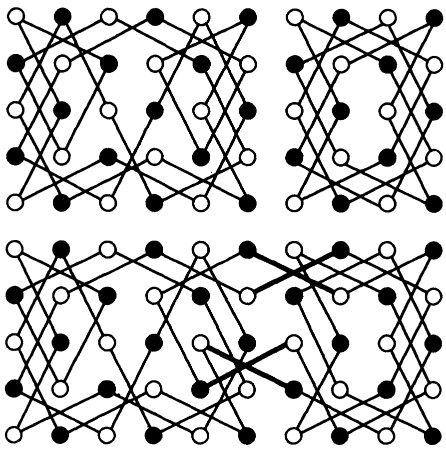


FIGURE 7
Extension of a Hamiltonian cycle in $G(5,n)$ to one in $G(5,n+4)$ for $n=6$.

Similarly, for m even as in FIGURE 8, we can incorporate $H(m,4)$ into a Hamiltonian cycle of $G(m,n)$ by first removing the four edges

$$(1,n-1)\text{–}(3,n),\ (1,n)\text{–}(3,n-1),\ (m-2,n-1)\text{–}(m,n),\ \text{and}\ (m-2,n)\text{–}(m,n-1)$$

from the Hamiltonian cycle, then removing the four edges

$$(2,n+1)\text{–}(4,n+2),\ (2,n+2)\text{–}(4,n+1),\ (m-3,n+1)\text{–}(m-1,n+2),\ \text{and}\ (m-3,n+2)\text{–}(m-1,n+1)$$

from $H(m,4)$, and finally inserting the eight edges

$$(1,n-1)\text{–}(2,n+1),\ (1,n)\text{–}(2,n+2),\ (3,n-1)\text{–}(4,n+1),\ (3,n)\text{–}(4,n+2),$$
$$(m-2,n-1)\text{–}(m-3,n+1),\ (m-2,n)\text{–}(m-3,n+2),$$
$$(m,n-1)\text{–}(m-1,n+1),\ \text{and}\ (m,n)\text{–}(m-1,n+2).$$

Again, the new cycle contains the prescribed 10 edges. The extension of $G(6,6)$ to $G(6,10)$ is illustrated in FIGURE 8. This completes the proof of the lemma.

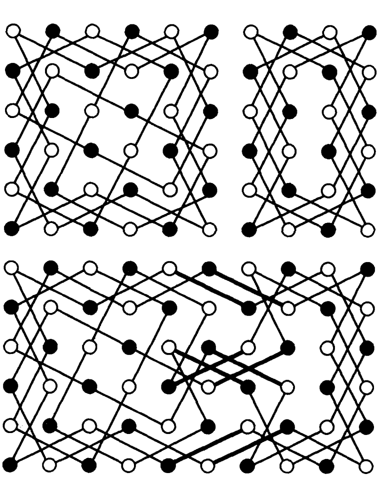


FIGURE 8
Extension of a Hamiltonian cycle in $G(6,n)$ to one in $G(6,n+4)$ for $n=6$.

But now we must complete the proof of the theorem. The lemma can be used to add four columns or four rows to a known solution. Thus, we can construct solutions for $G(m, n)$ provided we have a collection of starting cases for each possible modulo class pair $[i, j]$ where both i and j are taken modulo 4. Thus, it might seem that we would need 16 instances to serve as the base of our construction. But by flipping a rectangle over about its main diagonal we can interchange i and j, reducing the count to 10 instances. Moreover, recall that condition (a) forbids odd ordered boards, excluding classes [1, 1], [1, 3], and [3, 3]. This leaves just seven classes. Considering the forbidden values of $m = 1$, 2, and 4, the smallest possible members of these seven classes are 3×6, 3×8, 5×6, 5×8, 6×6, 6×8, and 8×8. But condition (c) excludes size 3×6. To replace it and be able to generate all other orders in the same class, we must include both 3×10 and 7×6. Similarly, the impossible order 3×8 forces us to include both 3×12 and 7×8. Thus, there are nine specific instances whose Hamiltonian cycles must be constructed individually before the lemma can be used to finish the job by induction. I have no particular method for generating these nine solutions, but whenever possible I have tried to select solutions that have pleasing symmetry or near symmetry. All nine are collected in FIGURE 9. I couldn't resist the urge to seek the most compact arrangement of the nine solutions into a single drawing.

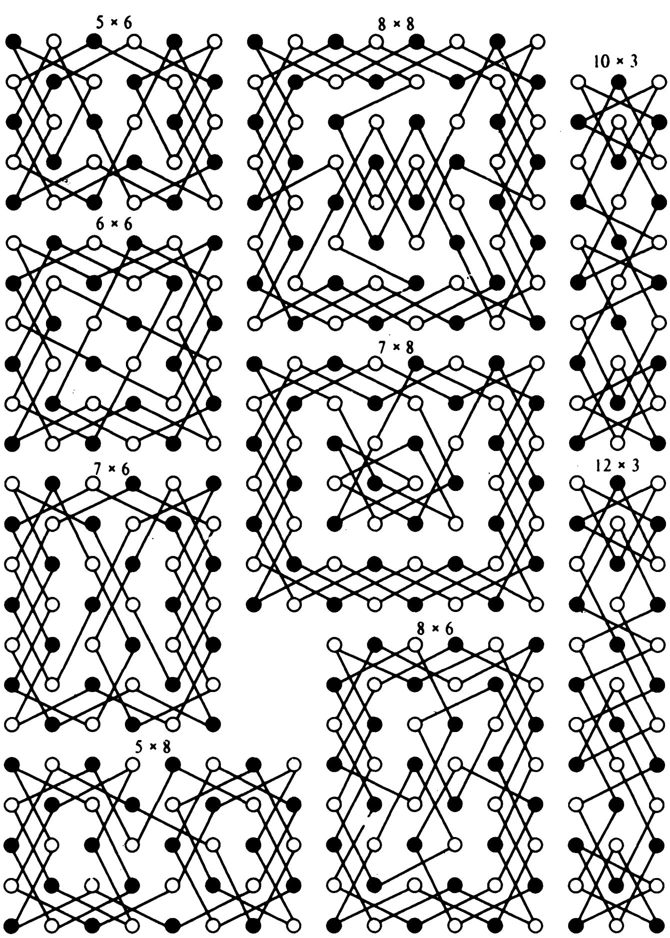


FIGURE 9

The nine Hamiltonian cycles that form the base of the inductive construction.

I like this illustration that induction can require many cases in order to get started because it shows students that they must be flexible when designing a proof by induction. I find this complete solution satisfying because the three conditions in the theorem are so easy to state, the impossible sizes are easy to understand, and while the inductive construction requires a bit of detail, the method remains totally elementary. Not everyone may wish to take class time to go through the complete solution, but students should be told that the full solution is entirely within their grasp.

Acknowledgement. The author is grateful to the referees for the detailed comments that greatly improved this article.

REFERENCES

1. C. Berge, *Graphs and Hypergraphs*, North-Holland, Amsterdam, 1973, p. 188.
2. G. Chartrand, *Introductory Graph Theory*, Dover Publications, Mineola, NY, 1985, pp. 133–135.
3. H. Dudeney, *Amusements in Mathematics*, Dover Publications, Mineola, NY, 1970, p. 103.
4. R. B. Eggleton and A. Eid, Knight's circuits and tours, *Ars Combinatoria* 17A (1984), 145–167.
5. R. Honsberger, Pósa's solution, *Mathematical Gems*, 1973, p. 145.
6. W. W. Rouse Ball and H. S. M. Coxeter, *Mathematical Recreations and Essays*, University of Toronto Press, Toronto, 1974, pp. 175–185.
7. R. J. Wilson and J. J. Watkins, *Graphs, an Introductory Approach*, John Wiley and Sons, Inc., New York, 1990, p. 145.

Steiner Minimal Trees on Chinese Checkerboards

F. K. HWANG
AT&T Bell Laboratories
Murray Hill, NJ 07974

D. Z. DU
Academia Sinica
Beijing, China 100010

1. Introduction For a given set P of points in the plane, $S(P)$, *a Steiner minimal tree* (SMT) is the shortest network spanning P. Recently, Chung, Gardner, and Graham [1] studied the SMT problem on an $n \times n$ square lattice that they called a *checkerboard*. They gave a general construction of heuristic SMTs on checkerboards and conjectured that the constructed trees are minimal for certain values of n. While the checkerboard is in the shape of a square, the Chinese checkerboard is in the shape of a Star of David triangulated into equilateral triangles of the same size (see FIGURE 1).

Formally, for $n \geq 2$ we define a latticed hexagon (triangle) to be a regular hexagon (equilateral triangle) divided into disjoint equilateral triangles by lines parallel to the sides of the hexagon (triangle). The hexagon (triangle) is said to be of order n if there are n lattice points on each side. Let $H_n(T_n)$ be the set of lattice points of the hexagon (triangle) of order n where $H_1 = T_1$ is defined to be a single point. A Chinese checkerboard of order n, denoted by C_n, consists of an H_n in the center and a T_n attached to each side of the H_n. We assume that all the lattice segments are of length one and we call each equilateral triangle of unit side a *basic triangle*.

In this paper we construct SMTs spanning H_n, T_n, and C_n, respectively, for all n. While the problem of constructing SMTs is NP-hard in general [3], we show that our constructions yield SMTs for (i) H_n for all n, (ii) T_n for $n \equiv 1$ or 2 (mod 4), (iii) C_n for all n.

2. The H_n construction Note that H_n contains a nest of concentric H_k for $k = 1, \ldots, n-1$ and every lattice point of H_n belongs to a side of one of these H_k, with H_1 consisting of a single point in the center. We will call the space between a concentric H_k and a concentric H_{k+1} the kth layer of H_n, denoted by L_k. It is easily verified that the sides of H_k contain a total of $6(k-1)$ lattice points for $k \geq 2$ and one lattice point for $k = 1$. For each $k = 2, \ldots, n$, order the $6(k-1)$ lattice points along a clockwise tour (starting from any point) into $p_{k,1}, \ldots, p_{k,6(k-1)}$. We now give an algorithm h that constructs a tree h_n, interconnecting all lattice points in H_n. Construct a $S(T_2)$ for each basic triangle in L_{k-1} containing the two points $p_{k,2i-1}$ and $p_{k,2i}$, $i = 1, \ldots, 3(k-1)$, $2 \leq k \leq n$. Then h_n is the union of all these $S(T_2)$. FIGURE 2 illustrates h_2, h_3, and h_4. (Since h_n can be constructed recursively, only these $S(T_2)$ in L_{n-1} are explicitly shown.)

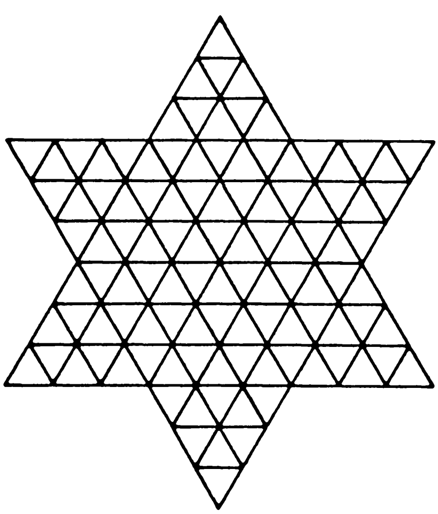


FIGURE 1
A Chinese checkerboard.

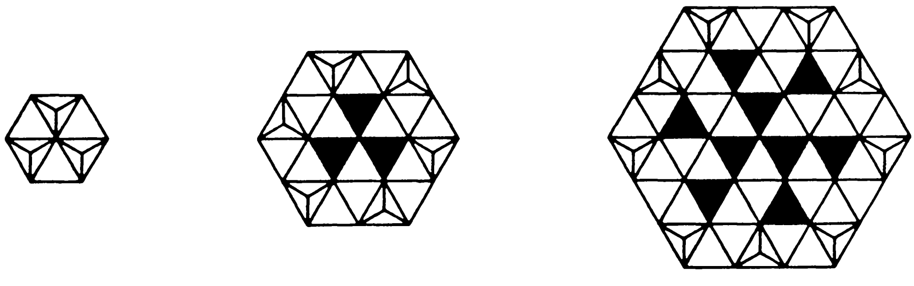


FIGURE 2
h_2, h_3, and h_4.

Let $p(N)$ denote the number of lattice points and $s(N)$ the number of $S(T_2)$ in a network N.

THEOREM 1. *h_n spans H_n and the length of $h_n = \ell(h_n) = 3\sqrt{3}\,n(n-1)/2$.*

Proof. We first prove the spanning part. Theorem 1 is trivially true for h_2. We prove the general case h_n by induction on n. Suppose that all lattice points in H_n that are also contained in the concentric H_{n-1} have been spanned by h_{n-1}. We need only be concerned with whether h_n connects the lattice points on the sides of H_n to h_{n-1}. But this is easily confirmed since each such point is connected to a point on a side of H_{n-1} through an $S(T_2)$ in L_{n-1}.

We next prove the length part. Note that h_n contains $3kS(T_2)$ in L_k. Hence

$$s(h_n) = \sum_{k=1}^{n-1} 3k = 3n(n-1)/2.$$

Therefore,

$$\ell(h_n) = s(h_n)\cdot\ell(S[T_n]) = 3\sqrt{3}\,n(n-1)/2.$$

3. The T_n construction Take a side a of T_n, say, the bottom side, and order its lattice points from left to right into $a_1,\dots,a_n$. Similarly, we define $b_1,\dots,b_n$ and $c_1,\dots,c_n$ for the two sides right and left, respectively, of the bottom side. For $n \geqslant 4$, define *peeling* to be the construction of $S(T_2)$ for every basic triangle containing p_i and p_{i+1}, $i = 1,\dots,n-3$, where p_i can represent a_i, b_i and c_i. Then it is easily seen that peeling creates three disjoint strips, each consisting of $n - 3S(T_2)$'s. We will call them strips a, b, and c. Furthermore, the points of T_n left unpeeled form a T_k where $k = 1$ for $n = 4$, $k = 0$ for $n = 5$, and $k = n - 6$ for $n \geq 6$ (see FIGURE 3).

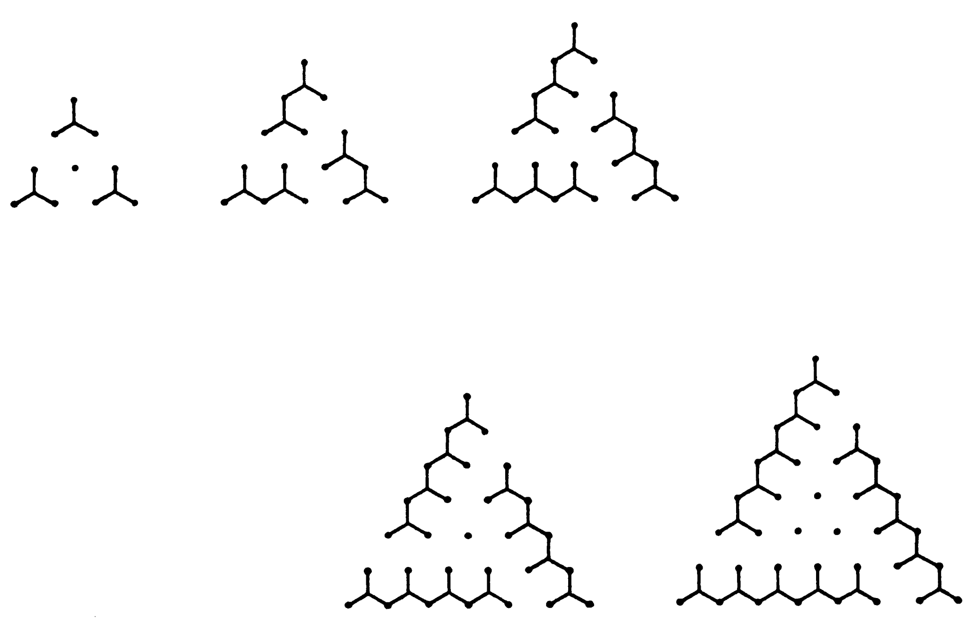


FIGURE 3
Peeling T_4, T_5, T_6, T_7, T_8.

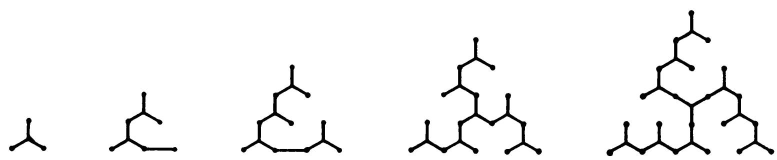


FIGURE 4
t_n for $1 \le n \le 6$.

We define an algorithm t that recursively constructs a tree t_n interconnecting T_n. FIGURE 4 gives t_n for $1 \le n \le 6$.

For $n \ge 7$ we peel T_n and construct a t_{n-6} for the unpeeled T_{n-6}. We consider two cases:

Case (i). $n \equiv 0$ or 3 (mod 4). It is easily verified that there exists a basic triangle Δxyz where x and y belong to strip a and strip c, and z is a tip of T_{n-6}. Construct an $S(T_2)$ for $\{x, y, z\}$ that connects strip a, strip c, and t_{n-6}. Also construct the edge $[a_{n-3}, a_{n-2}]$, called the *odd edge*, to connect strip a with strip b. Then t_n consists of the union of its peeling with t_{n-6}, with the $S(T_2)$ on $\{x, y, z\}$, and with the edge $[a_{n-3}, a_{n-2}]$. FIGURE 5 illustrates t_7 and t_8 constructed in this manner.

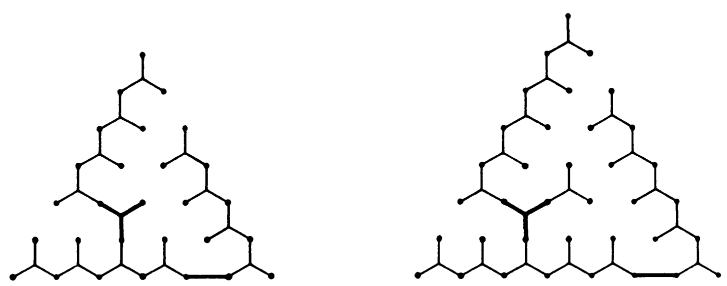


FIGURE 5
t_7 and t_8.

Case (ii). $n \equiv 1$ or 2 (mod 4). Peel T_n and construct $t_{n'} = t_{n-6}$. Delete from $t_{n'}$ the edge $[u, v]$ where $[u, v] = [a_{n'-3}, a_{n'-2}]$ if $n' \ge 7$ and $[u, v] = [a_2, a_3]$ if $n' = 3$ or 4. Construct an $S(T_2)$ for the basic triangle Δuvw where w is a lattice point on strip a of T_n. Also construct an $S(T_2)$ for a basic triangle Δxyz where x and y belong to strip b and strip c, respectively, of T_n, and z is a tip of T_{n-6}.

FIGURE 6 illustrates t_9 and t_{10} constructed in this manner.

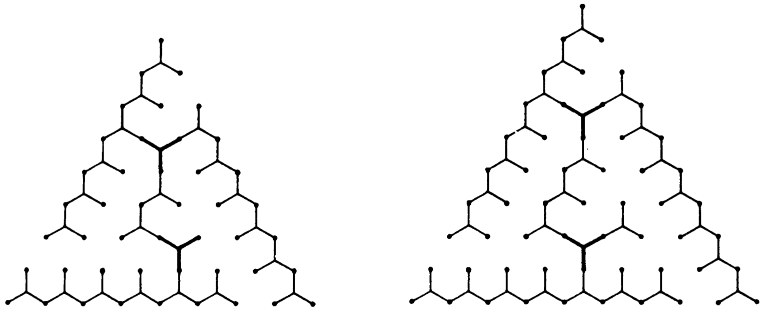


FIGURE 6
t_9 and t_{10}.

THEOREM 2. *t_n spans T_n and*

$$\ell(t_n) = \sqrt{3}\,(n+2)(n-1)/4 \qquad \text{for } n \equiv 1 \text{ or } 2 \pmod 4,$$
$$\ell(t_n) = 1 + \sqrt{3}\,[(n+1)n/4 - 1] \quad \text{for } n \equiv 0 \text{ or } 3 \pmod 4.$$

Proof. That t_n spans T_n is clear from its construction and using induction. For $n \geq 4$

$$s(t_n) = 3(n-3) + s(t_{n-6}) + 2 \quad \text{for } n \equiv 1 \text{ or } 2 \pmod 4,$$
$$s(t_n) = 3(n-3) + s(t_{n-6}) + 1 \quad \text{for } n \equiv 0 \text{ or } 3 \pmod 4.$$

Noting $s(t_1) = 0$, $s(t_2) = 1$, $s(t_3) = 2$, $s(t_4) = 4$, $s(t_5) = 7$, and $s(t_6) = 10$, we obtain

$$s(t_n) = (n+2)(n-1)/4 \quad \text{for } n \equiv 1 \text{ or } 2 \pmod 4,$$
$$s(t_n) = (n+1)n/4 - 1 \quad \text{for } n \equiv 0 \text{ or } 3 \pmod 4.$$

Theorem 2 now follows immediately.

4. The C_n construction For each component T_n of C_n we will designate the lattice point farthest from H_n as the tip of T_n. We define T_k^* to be the unique $T_k \leq T_n$, which contains the tip of T_n. We also designate *side a* to be the side of T_k^* opposite the tip of T_n. Define L_k to be the layer between side a of T_k^* and side a of T_{k+1}^*. Then C_n consists of the H_n and six T_{n-1}^*, each of which is attached to the H_n though its L_{n-1}.

We now define an algorithm c that constructs a tree c_n spanning C_n. We consider three cases:

Case (i). $n \equiv 0$ or $1 \pmod 4$. Construct a t_{n-1} for each T_{n-1}^* and construct a h_n for the H_n. Note that each t_{n-1} has an odd edge. We will call a basic triangle between a T_{n-1}^* and the H_n, with the odd edge as a side, an *odd triangle*. Connect each t_{n-1} with the h_n by substituting an $S(T_2)$ in the odd triangle for the odd edge (see FIGURE 7).

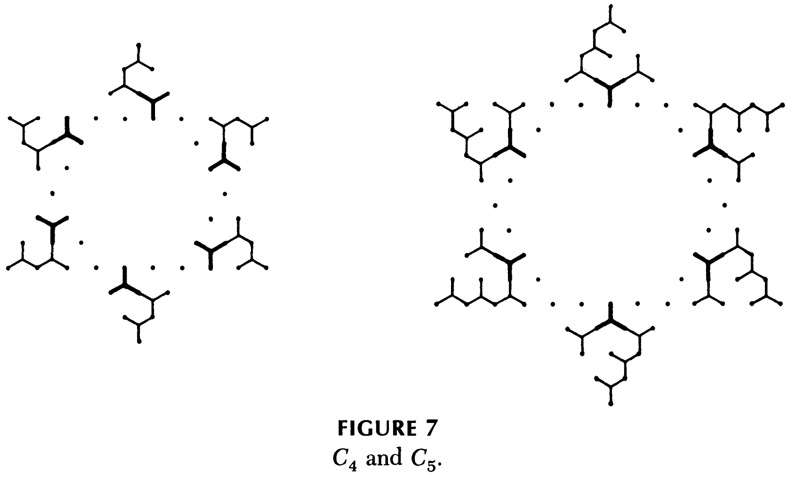
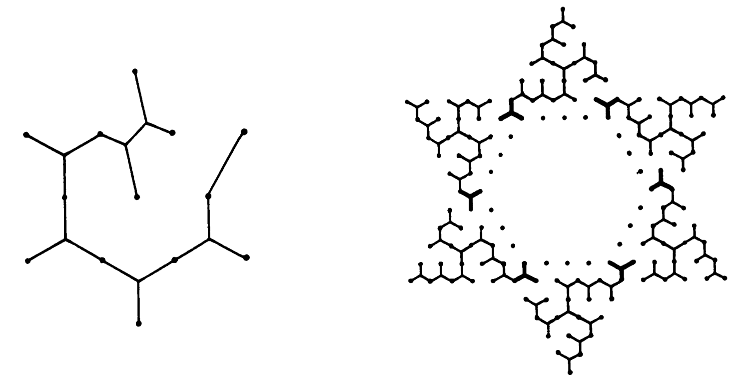


FIGURE 7
C_4 and C_5.

Case (ii). $n \equiv 2 \pmod 4$. For $n = 2$ c_2 is shown in FIGURE 8(a).

For $n > 2$ construct a h_{n-1} for the concentric H_{n-1} and construct a t_n for each T_n^* except flipping the $S(T_2)$ connecting the two rightmost points at side a inward to



(a) (b)

FIGURE 8
C_2 and C_6.

connect the h_{n-1} (see FIGURE 8(b)). Although these two points are no longer directly connected with their own t_n, they are connected to the t_n to their right since two neighboring T_n share a lattice point. The newly added $S(T_2)$ connects the right t_n to h_{n-1}, and through the latter to the original t_n.

Case (iii). $n \equiv 3 \pmod 4$. Construct a t_n for each T_n^* and an h_{n-1} for the concentric H_{n-1}. Again we note that each t_n has an odd edge. The deletion of the odd edge in a t_n separates strip b from strips a and c, but this strip b is connected to strip a of a neighboring t_n (called a *combination piece*) since they share a lattice point. Therefore, after the six odd edges are deleted, we have seven disjoint components, one for h_{n-1} and one each for the six combination pieces. By inserting an $S(T_2)$ in an odd triangle we connect to h_{n-1} not only the t_n whose odd edge defines this odd triangle but also the neighboring t_n that annexes strip b. Do this to every other odd triangle so that all seven components are connected (see FIGURE 9).

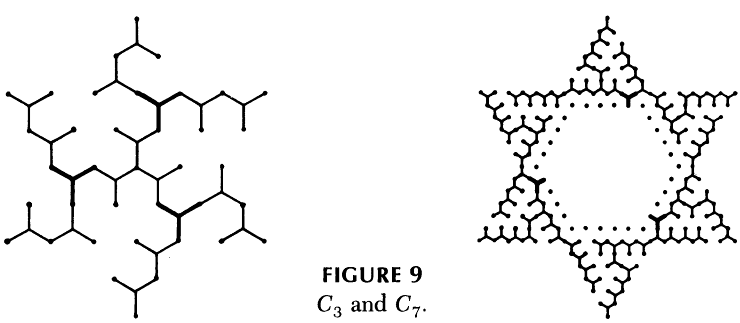


FIGURE 9
C_3 and C_7.

THEOREM 3. *c_n spans C_n and*

$$\ell(c_n) = 1 + \sqrt{7} + 4\sqrt{3} \quad \text{for } n = 2,$$
$$\ell(c_n) = 3\sqrt{3}\, n(n-1) \quad \text{for } n \neq 2.$$

Proof. That c_n spans C_n is clear from its construction. For $n \equiv 0$ or $1 \pmod 4$

$$\begin{aligned} s(c_n) &= s(h_n) + 6[s(t_{n-1}) + 1] \\ &= 3n(n-1)/2 + 6n(n-1)/4 \\ &= 3n(n-1). \end{aligned}$$

For $n \equiv 2 \pmod 4 \geqslant 6$

$$\begin{aligned} s(c_n) &= s(h_{n-1}) + 6s(t_n) \\ &= 3(n-1)(n-2)/2 + 6(n+2)(n-1)/4 \\ &= 3n(n-1). \end{aligned}$$

For $n \equiv 3 \pmod 4$

$$\begin{aligned} s(c_n) &= s(h_{n-1}) + 6t_n + 3 \\ &= 3(n-1)(n-2)/2 + 6[(n+1)n/4 - 1] + 3 \\ &= 3n(n-1). \end{aligned}$$

Thus $\ell(c_n) = 3\sqrt{3}\,n(n-1)$ for $n \neq 2$. It is straightforward to verify for $\ell(c_2)$. The $\sqrt{7}$ comes from a 4-point Steiner minimal tree [2].

5. An optimality analysis For a given set P of points, a hexagonal tree for P is a tree spanning P using lines of three given directions with 120° subtending angles. Let $W_D[P]$ denote the minimal hexagonal tree with D being the set of directions. Weng [4] observed that a hexagonal tree (with given D) can always be constructed from a Steiner tree by substituting two hexagonal edges for each edge in the Steiner tree. Such a substitution can increase the length by a factor of at most $2/\sqrt{3}$ (which occurs when the two hexagonal edges are of equal length). Hence

LEMMA. *$\ell(S[P])/\ell(W_D[P]) \geq \sqrt{3}/2$ for any D and P.*

COROLLARY. *Let t be a tree interconnecting P. If there exists a D such that $\ell(t)/\ell(W_D[P]) = \sqrt{3}/2$, then t is an* SMT.

We now use the above Corollary to show the optimality of our constructions.

THEOREM 4. *H_n for all n, T_n for $n \equiv 1$ or $2 \pmod 4$, and C_n for all n are* SMTs.

Proof. Let D consist of the three directions of the concerned lattice. Then clearly, W_D is the minimal spanning tree whose length is the number of lattice points minus one. It is easily verified that

$$\begin{aligned} p(H_n) &= 1 + 3n(n-1), \\ p(T_n) &= n(n+1)/2, \end{aligned}$$

and

$$p(C_n) = 1 + 6n(n-1)$$

Using Theorems 1, 2, and 3, we have

$$\begin{aligned} \ell(h_n)/\ell(W_D[H_n]) &= [3\sqrt{3}\,n(n-1)/2]/[3n(n-1)] = \sqrt{3}/2, \text{ for all } n, \\ \ell(t_n)/\ell(W_D[T_n]) &= [\sqrt{3}\,(n+2)(n-1)/4]/[n(n+1)/2 - 1] \\ &= \sqrt{3}/2 \qquad \text{for } n \equiv 1 \text{ or } 2 \pmod 4, \end{aligned}$$

and

$$\ell(c_n)/\ell(W_D[C_n]) = [3\sqrt{3}\,n(n-1)]/[6n(n-1)] = \sqrt{3}/2 \qquad \text{for } n \neq 2.$$

Theorem 4 follows from the Corollary immediately except for C_2. That C_2 is an SMT has been verified by W. D. Smith with computer aid [5].

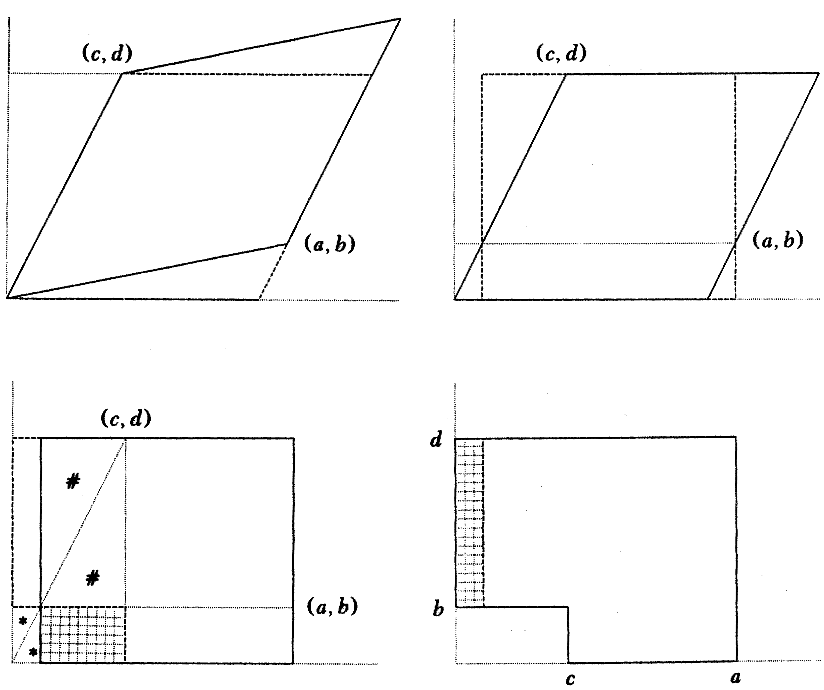
Acknowledgment. The authors wish to thank D. S. Johnson for suggesting a simpler construction of C_n for $n \equiv 0$ or 1 (mod 4).

REFERENCES

1. F. R. K. Chung, M. Gardner, and R. L. Graham, Steiner trees on a checkerboard, this MAGAZINE 62 (1989), 83–96.
2. E. N. Gilbert and H. O. Pollak, Steiner minimal trees, *SIAM J. Appl. Math.* 16 (1968), 1–29.
3. M. R. Garey, R. L. Graham, and D. S. Johnson, The complexity of computing Steiner minimal trees, *SIAM J. Appl. Math.* 32 (1977), 835–859.
4. J. F. Weng, Steiner problem in hexagonal metric, unpublished manuscript.
5. W. D. Smith, personal communication.

Proof without Words:

Area of the Parallelogram Determined by Vectors (a, b) and $(c, d) = \pm\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \pm(ad - bc)$



—YIHNAN DAVID GAU
CALIFORNIA STATE UNIVERSITY
LONG BEACH, CA 90840

From Fences to Hyperboxes and Back Again

WILLIAM P. COOKE
West Texas State University
Canyon, TX 79016

To enclose maximum area of a rectangular field at fixed cost, allocate half the cost to each pair of opposite sides. To achieve maximum volume of a box at fixed cost, allocate one third the cost to each pair of opposite faces.

Did you know that? If not, then quite possibly you and I were the only two people left who didn't. If you did, then still you might find something of interest here. Specifically, you will see how the assertions in the first paragraph are only special cases of an n-dimensional generalization.

Consider the ordinary "fence problem" from beginning calculus: Maximize V, where

$$V = x_1 x_2,$$

subject to

$$a_1 x_2 + a_2 x_1 = C, \qquad x_1 > 0, \quad x_2 > 0. \tag{1}$$

Here x_1 and x_2 can be the dimensions of a rectangular field, V its area, and a_1 the *sum* of costs per unit length of fencing for the two opposite sides of length x_2 (a_2 is similarly interpretable). Thusly provision is made for a river or existing fence along one side of the field (assume zero cost per unit length for that side). C (positive) can be the fixed cost.

If x_1^* and x_2^* are the optimal dimensions yielding maximum $V, V_{\max}$, then an elementary calculus argument gives

$$x_1^* = a_1[C/(2a_1a_2)]^{1/1}, \qquad x_2^* = a_2[C/(2a_1a_2)]^{1/1}, \qquad V_{\max} = (C/2)[C/(2a_1a_2)]^{1/1}. \tag{2}$$

The rationale for the weird format in (2) will become clear as work progresses. More to the point for now, note that the cost allocated to the x_1-sides is $a_2x_1^* = C/2$, and that the cost allocated to the x_2-sides is also $C/2$.

The usual box problem (either open or closed) yields analogous results. Consider: Maximize V, where

$$V = x_1 x_2 x_3,$$

subject to

$$a_1x_2x_3 + a_2x_1x_3 + a_3x_1x_2 = C, \qquad x_1 > 0, \quad x_2 > 0, \quad x_3 > 0. \tag{3}$$

Here a_j, $j = 1, 2, 3$, may be interpreted as the sum of costs per unit area for the faces of the box for which x_j is not a facial dimension. A bit of work with partial derivatives gives as a solution for (3):

$$x_j^* = a_j[C/(3a_1a_2a_3)]^{1/2}, \qquad j = 1,2,3,$$
$$V_{\max} = (C/3)[C/(3a_1a_2a_3)]^{1/2}. \tag{4}$$

Given (4) it is no great task to show that $a_1x_2^*x_3^* = a_2x_1^*x_3^* = a_3x_1^*x_2^* = C/3$, which is what we wanted to observe.

By now you have anticipated the generalization. The n-dimensional "hyperbox" problem you perceive is: Maximize V, where

$$V = \prod_{i=1}^{n} x_i,$$

subject to

$$\sum_{j=1}^{n} a_j x_j^{-1} \prod_{i=1}^{n} x_i = C,\ n \geqslant 2,$$
$$x_j > 0, \qquad j = 1,2,\dots,n. \tag{5}$$

You have guessed its solution to be

$$x_j^* = a_j\left[C\Big/\left(n\prod_{i=1}^{n} a_i\right)\right]^{1/(n-1)}, \qquad j = 1,2,\dots,n,$$
$$V_{\max} = (C/n)\left[C\Big/\left(n\prod_{i=1}^{n} a_i\right)\right]^{1/(n-1)}, \tag{6}$$

accompanied by

$$a_j x_j^{*-1} \prod_{i=1}^{n} x_i^* = C/n, \qquad j = 1,2,\dots,n. \tag{7}$$

Are equations (7) really true? Do we maximize the "volume" of our n-dimensional box by simply allocating a cost of C/n to each set of corresponding flats? Yes.

Note that optimal dimensions and $V_{\max}$, given in (6), are not of primary interest here, but that (6) is implied by (7) and a little algebra. A very pretty, direct verification of (7) can be achieved by solving a re-phrased version of (5) as a zero degree of difficulty [1, p. 83] geometric programming problem. Geometric programming (see [2, pp. 220–241] for the essentials) will conclude that each term in the constraint in (5) is of equal importance to the design of the box, implying the cost allocation of C/n.

Most readers, however, will be more comfortable with calculus than with geometric programming. Here a Lagrange-multiplier argument [3, p. 199], that bypasses the optimal dimensions in (6), will be used to verify conjecture (7).

Consider the function F with value

$$F(x_1,\dots,x_n,\lambda) = \prod_{i=1}^{n} x_i - \lambda\left[\sum_{j=1}^{n} a_j x_j^{-1} \prod_{i=1}^{n} x_i - C\right]. \tag{8}$$

Letting asterisks denote optimal values of decision variables and setting the $n+1$ partial derivatives of F equal to zero gives the system

$$x_j^{*-1}\left\{\prod_{i=1}^{n} x_i^* - \lambda^*\left[\sum_{k=1}^{n} a_k x_k^{*-1}\prod_{i=1}^{n} x_i^* - a_j x_j^{*-1}\prod_{i=1}^{n} x_i^*\right]\right\} = 0, \qquad j = 1,2,\ldots,n, \tag{9}$$

and

$$\sum_{j=1}^{n} a_j x_j^{*-1}\prod_{i=1}^{n} x_i^* - C = 0, \tag{10}$$

with (9) coming from $\partial F/\partial x_j$, $j = 1,2,\ldots,n$ and (10) from $\partial F/\partial\lambda$.

Substitute C, from (10), for $\sum_{k=1}^{n} a_k x_k^{*-1}\prod_{i=1}^{n} x_i^*$ in (9), and multiply (9) by x_j^*. This gives

$$\prod_{i=1}^{n} x_i^* - \lambda^*\left[C - a_j x_j^{*-1}\prod_{i=1}^{n} x_i^*\right] = 0, \qquad j = 1,2,\ldots,n. \tag{11}$$

From (11) we observe that

$$a_j x_j^{*-1}\prod_{i=1}^{n} x_i^* = C - (1/\lambda^*)\prod_{i=1}^{n} x_i^*, \qquad j = 1,2,\ldots,n, \tag{12}$$

or that $a_j x_j^{*-1}\prod_{i=1}^{n} x_i^*$ is a constant, say K, for all j. Then the constraint in (5) becomes, at optimality,

$$\sum_{j=1}^{n} K = nK = C, \tag{13}$$

or

$$K = a_j x_j^{*-1}\prod_{i=1}^{n} x_i^* = C/n, \qquad j = 1,2,\ldots,n, \tag{14}$$

which was to be shown.

So we see that the basic structure of generalization (5) explains the cost allocations observed earlier for fences and boxes. When we come back again to the solutions of problems (1) and (3) we perceive those allocations as simply (14) when $n = 2$ and $n = 3$.

REFERENCES

1. R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming*, John Wiley and Sons, Inc., New York, 1967.
2. C. McMillan, Jr., *Mathematical Programming*, John Wiley and Sons, Inc., New York, 1970.
3. A. E. Taylor, *Advanced Calculus*, Ginn, New York, 1955.

An Interesting Subset of the Highly Composite Numbers

STEVEN RATERING
Central College
Pella, IA 50219

Several years ago I was dissatisfied with a common idea of what makes a number round. I thought that most people would think that 30, 200, and 8000 are rounder than 42, 147, or 1372, for example. However, if we write these numbers in base 7, the first three are 42, 404, and 32216, while the last three are 60, 300, and 4000. In base 7, the latter three look rounder. I wanted a definition of round that didn't depend on the representation of a number. It seems that numbers with many divisors should be thought of as rounder than numbers with few divisors. The size of a number should also be taken into consideration—smaller numbers can't be expected to have as many divisors as larger numbers.

I settled on the following definition: A number is *round* if it has more divisors than all numbers less than it. The first few numbers that have this property are 1, 2, 4, 6, 12, and 24. Notice that most of these numbers are divisors of all the larger ones. If we look further in the list, however, the pattern changes. The next few round numbers are 36, 48, 60, 120, 180, 240, 360, 720, and 840. Among these round numbers, only 1, 2, 6, 12, and 60 have the property that they are divisors of all the larger numbers in the list. This leads to another definition: A number is *really round* if it is round and is a divisor of all larger round numbers. After looking even further in the list, I found only one more really round number, 2520, and it didn't seem likely that there were any more. This leads to the conjecture: There are exactly six really round numbers, 1, 2, 6, 12, 60, and 2520. I recently found out that Ramanujan studied round numbers [**2**], but called them highly composite numbers (HCNs), even though 1 and 2 have this property but are not even composite. I adopted his terminology and changed "really round number" to "special highly composite number" (SHCN). When I received a paper on HCN's by J. L. Nicolas [**1**], I glanced through it quickly and saw the "SHCNs" in print. I was pleased that someone else had found them worthy of study and wondered if Nicolas' proof was similar to mine. But a closer inspection showed that Nicolas' "SHCNs" were *superior* highly composite numbers, and though they include the special highly composite numbers, there are infinitely many of them and their definition is messier. A number, n, is a superior HCN if there exists an $e>0$ such that for all positive integers m, $d(n)/n^e \geqslant d(m)/m^e$, when $d(n)$ equals the number of divisors of n. Without further ado, here are some formal definitions and proofs.

Definitions:

A *Highly Composite Number* is a number (positive integer) that has more divisors than all lesser numbers.

A *Special Highly Composite Number* is an HCN that is a divisor of all greater HCNs.

I shall prove that there are exactly six SHCNs, namely, 1, 2, 6, 12, 60, and 2520.

The first 18 HCNs together with their prime factorizations (bases omitted) and number of divisors are:

1	—				1
2	1				2
4	2				3
6	1	1			4
12	2	1			6
24	3	1			8
36	2	2			9
48	4	1			10
60	2	1	1		12
120	3	1	1		16
180	2	2	1		18
240	4	1	1		20
360	3	2	1		24
720	4	2	1		30
840	3	1	1	1	32
1260	2	2	1	1	36
1680	4	1	1	1	40
2520	3	2	1	1	48

The divisors of $n = 2^{a_1} \cdot 3^{a_2} \cdot 5^{a_3} \cdots p_k^{a_k}$, where p_k is the kth prime and $a_i \geqslant 0$ for $i = 1, \ldots, k$, are of the form $2^{b_1} \cdot 3^{b_2} \cdot 5^{b_3} \ldots p_k^{b_k}$, where $0 \leqslant b_i \leqslant a_i$ for $i = 1, \ldots, k$. There are $a_1 + 1$ choices for the value of b_1, $a_2 + 1$ choices for the value of b_2, and so on, so $d(n) = (a_1 + 1)(a_2 + 1) \ldots (a_k + 1)$.

Therefore, $d(n)$ depends only on the exponents of the prime factorization of n and not on the bases.

LEMMA 1. *All HCNs are of the form* $2^{a_1} \cdot 3^{a_2} \cdots p_k^{a_k}$, *where* $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_k$.

Proof. Suppose not; then there exists an HCN, n, of the form $2^{a_1} \cdot 3^{a_2} \cdots p_k^{a_k}$, where $a_i < a_j$ and $1 \leqslant i < j \leqslant k$. But then $n' = np_i^{a_j - a_i} p_j^{a_i - a_j}$ has as many divisors as n, but $n' < n$ so n is not an HCN, which is a contradiction, and the lemma is proved.

LEMMA 2. 1 *and* 2 *are SHCNs.*

Proof. Clear.

THEOREM 3. 6 *is an SHCN.*

Proof. Suppose not; then there exists an HCN, $n > 6$, and 6 does not divide n, but by lemma 2 and the definition of an SHCN, 2 does divide n. Therefore, 3 does not divide n. This, together with lemma 1, implies that $n = 2^k$ for some $k \geqslant 3$. But $2^{k-2} \cdot 3 < 2^k$ and 2^k is an HCN, so $d(2^{k-2} \cdot 3) < d(2^k)$, implying $2(k-1) < k + 1$. Thus $k < 3$, which is a contradiction, so 6 is an SHCN.

THEOREM 4. 12 *is an SHCN.*

Proof. Suppose not; then there exists an HCN, $n > 12$, and 12 does not divide n, but 6 does, by theorem 3. Therefore, 2's exponent in n's prime factorization is a one, and then by lemma 1 and the fact that n is larger than 12, we know that $n = 2 \cdot 3 \cdot 5 \cdot r$ where r is a (possibly empty) product of distinct primes greater than 5. Since $2^3 \cdot 3 \cdot r < 2 \cdot 3 \cdot 5 \cdot r$ and $2 \cdot 3 \cdot 5 \cdot r$ is an HCN, $d(2^3 \cdot 3 \cdot r) < d(2 \cdot 3 \cdot 5 \cdot r)$. This implies $8 \cdot d(r) < 8 \cdot d(r)$, which is a contradiction, so 12 is an SHCN.

THEOREM 5. *60 is an SHCN.*

Proof. Suppose not; then there exists an HCN, $n > 60$, and 60 does not divide n, but 12 does, by theorem 4. This means that 3 divides n, but 5 does not divide n, and then by lemma 1, we can let $n = 2^a \cdot 3^b$, where $a \geqslant b \geqslant 1$.

Case I. $b = 1$.
Then $a \geqslant 5$ because $2^a \cdot 3 > 60$. Since $2^{a-3} \cdot 3 \cdot 5 < 2^a \cdot 3$ and $2^a \cdot 3$ is an HCN, $d(2^{a-3} \cdot 3 \cdot 5) < d(2^a \cdot 3)$. Thus $4(a-2) < 2(a+1)$, which implies that $a < 5$, which is a contradiction.

Case II. $b = 2$.
Then $a \geqslant 3$ because $2^a \cdot 3^2 > 60$. Since $2^{a-1} \cdot 3 \cdot 5 < 2^a \cdot 3^2$ and $2^a \cdot 3^2$ is an HCN, $d(2^{a-1} \cdot 3 \cdot 5) < d(2^a \cdot 3^2)$. Thus $4a < 3(a+1)$, which implies that $a < 3$, which is a contradiction.

Case III. $b \geqslant 3$.
Then $a \geqslant 3$ because $a \geqslant b$. Since $2^{a-1} \cdot 3^{b-1} \cdot 5 < 2^a \cdot 3^b$ and $2^a \cdot 3^b$ is an HCN, $d(2^{a-1} \cdot 3^{b-1} \cdot 5) < d(2^a \cdot 3^b)$. Thus $2ab < (a+1)(b+1)$, which implies that $ab < a + b + 1$. Since $b \geqslant 3$, $3a \leqslant ab < a + b + 1$, and $2a < b + 1$. Again using $a \geqslant b$, we have $2b \leqslant 2a < b + 1$, and $b < 1$, which is a contradiction.

Since all cases lead to contradictions, the theorem is proved.

THEOREM 6. *$2520 = 2^3 \cdot 3^2 \cdot 5^1 \cdot 7^1$ is an SHCN.*

Proof. Suppose not; then there exists an HCN, $n > 2520$, and 2520 does not divide n, but $60 = 2^2 \cdot 3^1 \cdot 5^1$ does, by theorem 5. Therefore 4, 3, and 5 are factors of n, but at least one of 8, 9, and 7 is not.

Case I. 7 does not divide n. By lemma 1 we can let $n = 2^a \cdot 3^b \cdot 5^c$.

If $a \geqslant 3$ then $2^{a-3} \cdot 3^b \cdot 5^c \cdot 7 < 2^a \cdot 3^b \cdot 5^c$, and since $2^a \cdot 3^b \cdot 5^c$ is an HCN, $2(a-2) < (a+1)$, and $a < 5$; so $a < 5$ in any case.

If $b \geqslant 2$, then $2^a \cdot 3^{b-2} \cdot 5^c \cdot 7 < 2^a \cdot 3^b \cdot 5^c$, and since $2^a \cdot 3^b \cdot 5^c$ is an HCN, $2(b-1) < (b+1)$, and $b < 3$; so $b < 3$ in any case.

But $c \leqslant b$, so $c < 3$ as well.

We have upper bounds on a, b, and c, and since $2^a \cdot 3^b \cdot 5^c > 2520$, we must have $a = 4$, $b = 2$, and $c = 2$. Then $n = 3600$ and since n is an HCN, $d(2520) < d(3600)$. But then $48 < 45$, which leads to a contradiction.

Case II. 9 does not divide n, but 7 does. By lemma 1 we can let $n = 2^a \cdot 3^1 \cdot 5^1 \cdot 7^1 \cdot 11^e \cdot r$ where $e = 0$ or 1, and r has no divisors among 2, 3, 5, 7, and 11.

If $e = 1$, then $2^a \cdot 3^3 \cdot 5^1 \cdot 7^1 \cdot r < 2^a \cdot 3^1 \cdot 5^1 \cdot 7^1 \cdot 11^1 \cdot r$ and since n is an HCN, $4 < 2 \cdot 2$, but that is not the case, so $e = 0$ and $r = 1$. Thus $n = 2^a \cdot 3^1 \cdot 5^1 \cdot 7^1$. Since $n > 2520, a \geqslant 5$. But $2^{a-2} \cdot 3^2 \cdot 5^1 \cdot 7^1 < 2^a \cdot 3^1 \cdot 5^1 \cdot 7^1$, and since n is an HCN, $3(a-1) < 2(a+1)$ and $a < 5$, which is a contradiction.

Case III. 8 does not divide n, but 7 and 9 do. By lemma 1 we can let $n = 2^2 \cdot 3^2 \cdot 5^c \cdot 7^d \cdot 11^e \cdot r$ where r has no divisors among 2, 3, 5, 7, and 11.

If $e > 0$ then $2^5 \cdot 3^2 \cdot 5^c \cdot 7^d \cdot 11^{e-1} \cdot r < 2^2 \cdot 3^2 \cdot 5^c \cdot 7^d \cdot 11^e \cdot r$ and since n is an HCN, $6e < 3(e+1)$, and $e < 1$, which is a contradiction, so $e = 0$ and $r = 1$. Thus, $n = 2^2 \cdot 3^2 \cdot 5^c \cdot 7^d$.

If $c = 2$, then $2^4 \cdot 3^2 \cdot 5^1 \cdot 7^d < 2^2 \cdot 3^2 \cdot 5^2 \cdot 7^d$ and since n is an HCN, $5 \cdot 2 < 3 \cdot 3$, which is not the case, so $c = 1$ and $d = 1$. Thus $n = 2^2 \cdot 3^2 \cdot 5^1 \cdot 7^1 < 2520$, which is a contradiction.

Since all cases lead to contradictions, the theorem is proved.

LEMMA 7. *There are no SHCNs less than* 2520 *other than* 1, 2, 6, 12, *and* 60.

Proof. See the list of the first 18 HCNs.

THEOREM 8. *The ratio between two consecutive HCNs both greater than* 2520 *is less than* 2.

Proof. Suppose not; then there exists a consecutive pair of HCNs, n and m, $2520 < n < m$ and $m/n \geqslant 2$. Since $d(2n) > d(n)$, we must have $m \leqslant 2n$; so $m = 2n$.

Since 2520 is an SHCN and n is an HCN > 2520, we can let $n = 2^a \cdot 3^b \cdot 5^c \cdot 7^d \cdot 11^e \cdot 13^f \cdot r$ where $a \geqslant 3$, $b \geqslant 2$, $c \geqslant 1$, $d \geqslant 1$, and r has no divisors among 2, 3, 5, 7, 11, and 13.

Since $2n$ and n are consecutive HCNs, $d(3n/2) \leqslant d(n)$ and $d(4n/3) \leqslant d(n)$. Thus $a(b+2) \leqslant (a+1)(b+1)$ and $(a+3)b < (a+1)(b+1)$, which imply $a \leqslant b+1$ and $2b - 1 \leqslant a$. Thus, $2b - 1 \leqslant b + 1$ and $b \leqslant 2$. Since $a \leqslant b+1$, $a \leqslant 3$. We previously had $a \geqslant 3$ and $b \geqslant 2$, so now we can say $a = 3$ and $b = 2$.

If $f \geqslant 1$, then $2^5 \cdot 3^3 \cdot 5^c \cdot 7^d \cdot 11^e \cdot 13^{f-1} \cdot r < 2^3 \cdot 3^2 \cdot 5^c \cdot 7^d \cdot 11^e \cdot 13^f \cdot r$ and since n is an HCN, $6 \cdot 4 \cdot f < 4 \cdot 3 \cdot (f+1)$ and $f < 1$; so $f = 0$ and $r = 1$. Thus, $n = 2^3 \cdot 3^2 \cdot 5^c \cdot 7^d \cdot 11^e$, where $c = 1$ or 2.

Suppose $c = 2$. Since $2^5 \cdot 3^2 \cdot 5^1 \cdot 7^d \cdot 11^e < 2^3 \cdot 3^2 \cdot 5^2 \cdot 7^d \cdot 11^e$ and since n is an HCN, $6 \cdot 2 < 4 \cdot 3$, which is not the case; so $c = 1$. Since $n > 2520$ and 2520 is an SHCN, $d = e = 1$. Thus $n = 27720$ and $d(n) = 96$. But since $d(45360) = 100$ and $45360 < m = 2n$, n and m are not consecutive HCNs, a contradiction. So the ratio between consecutive HCNs greater than 2520 is less than 2.

COROLLARY 9. *There are no SHCNs greater than* 2520, *the* 18*th HCN.*

Proof. For $n > 18$ the ratio of the $(n+1)$st HCN to the nth HCN is less than 2; so the nth HCN is not a divisor of the $(n+1)$st HCN, proving that the nth HCN is not an SHCN.

In summation, there are exactly six SHCNs, namely, 1, 2, 6, 12, 60, and 2520.

In closing, I leave you with a conjecture and a problem. The conjecture: Each HCN other than 1 is the product of an HCN and a prime.* The problem: Write a computer program that can calculate the first 1000 HCNs in a reasonable amount of time.

Finally, I would like to express my appreciation to Agnes Andreassian for her helpful comments and to Bruce Berndt for sending me J. L. Nicolas' paper.

REFERENCES

1. J. L. Nicolas, On highly composite numbers. *Ramanujan Revisited*, Academic Press, 1988, pp. 215–244.
2. S. Ramanujan, Highly composite numbers, *Proc. London Math. Soc.* (2) 14 (1915), 347–409.

*Editor's Note: Several months after this note was accepted for publication the author found that this conjecture is false. Details of the counterexample may be obtained by writing the author.

A Fundamental Theorem of Calculus that Applies to All Riemann Integrable Functions

MICHAEL W. BOTSKO
Saint Vincent College
Latrobe, PA 15650

The usual form of the Fundamental Theorem of Calculus is as follows:

THEOREM 1. *Let f be Riemann integrable on* $[a, b]$ *and let g be a function such that* $g'(x) = f(x)$ *on* $[a, b]$. *Then*

$$\int_a^b f(x)\,dx = g(b) - g(a).$$

Unfortunately, this theorem only applies to Riemann integrable functions that are derivatives. Thus it cannot even be used to integrate the following simple function

$$f(x) = \begin{cases} 0 & \text{if } -1 \leqslant x < 0 \\ 1 & \text{if } 0 \leqslant x \leqslant 1. \end{cases}$$

It is the purpose of this note to present a theorem that does apply to every integrable function. In stating our result we will need the following definitions.

Definition 1. The function $f\colon [a, b] \to R$ satisfies a Lipschitz condition if there exists $M > 0$ such that

$$|f(x) - f(y)| \leqslant M|x - y| \quad \text{for all } x \text{ and } y \text{ in } [a, b].$$

Definition 2. A set E of real numbers has measure zero if for each $\varepsilon > 0$ there is a finite or infinite sequence $\{I_n\}$ of open intervals covering E and satisfying $\Sigma_n |I_n| \leqslant \varepsilon$ where $|I_n|$ is the length of I_n. If a property holds *except* on a set of measure zero, it is said to hold almost everywhere.

In [2] the author gave an elementary proof of the following result.

LEMMA. *If* $f\colon [a, b] \to R$ *satisfies a Lipschitz condition and* $f'(x) = 0$ *except on a set of measure zero, then f is a constant function on* $[a, b]$.

The proof required no measure theory other than the definition of a set of measure zero. This lemma was then used to prove that a bounded function that is continuous almost everywhere is Riemann integrable. We will use it here to establish our general form of the Fundamental Theorem of Calculus.

THEOREM 2. *Let f be Riemann integrable on* $[a, b]$ *and let g be a function that satisfies a Lipschitz condition and for which* $g'(x) = f(x)$ *almost everywhere. Then*

$$\int_a^b f(x)\,dx = g(b) - g(a).$$

Proof. Let $F(x) = \int_a^x f(t)\,dt$. Since f is bounded, F satisfies a Lipschitz condition. From the fact that f is continuous except on a set of measure zero (see [3] for an elementary proof), it follows that $F'(x) = f(x)$ almost everywhere. (This shows that every Riemann integrable function is almost everywhere the derivative of a function satisfying a Lipschitz condition.) It follows at once that

$$(F-g)'(x)=F'(x)-g'(x)=f(x)-f(x)=0$$

almost everywhere. In addition $F-g$ satisfies a Lipschitz condition. By the lemma there exists a real number k such that $F(x)=g(x)+k$ on $[a,b]$. Setting $x=a$ we have $k=-g(a)$. Finally, setting $x=b$, we get

$$\int_a^b f(x)\,dx=F(b)=g(b)-g(a),$$

which completes the proof.

Note that Theorem 2 includes Theorem 1 since any function that has a bounded derivative satisfies a Lipschitz condition.

Let us now integrate the following function. Define

$$f(x)=\begin{cases}-x & \text{if } x\in S=\{1,1/2,1/3,\dots\}\\ x^2+1 & \text{if } x\in[0,1]\setminus S.\end{cases}$$

Since f is bounded and continuous except on $S\cup\{0\}$, a set of measure zero, it is Riemann integrable. Let $g(x)=x^3/3+x$. Then g satisfies a Lipschitz condition and we have that $g'(x)=x^2+1=f(x)$ almost everywhere. Therefore,

$$\int_0^1 f(x)\,dx=g(1)-g(0)=4/3.$$

In this case $g'(x)\neq f(x)$ on an infinite set and yet Theorem 2 can still be used.

In closing, we give a useful corollary of Theorem 2.

COROLLARY. *Let f be Riemann integrable on $[a,b]$ and let g be a continuous function such that $g'(x)=f(x)$ except on a countable set. Then*

$$\int_a^b f(x)\,dx=g(b)-g(a).$$

Proof. To use Theorem 2 we need only show that g satisfies a Lipschitz condition. Since f is integrable there exists $M>0$ such that $|f(x)|\leqslant M$ for all x in $[a,b]$. Thus $-M\leqslant g'(x)\leqslant M$ except on a countable subset of $[a,b]$. Let $h(x)=Mx-g(x)$. Since h is continuous on $[a,b]$ and $h'(x)=M-g'(x)\geqslant 0$ except on a countable set, it follows from a result in [1] that h is increasing on $[a,b]$. Thus for c and d in $[a,b]$ with $c<d$ we have $h(c)\leqslant h(d)$ which gives $g(d)-g(c)\leqslant M(d-c)$. Similarly, we can show that $-M(d-c)\leqslant g(d)-g(c)$ and therefore $|g(d)-g(c)|\leqslant M(d-c)$. Thus g satisfies a Lipschitz condition and the proof follows immediately from Theorem 2.

REFERENCES

1. R. P. Boas, *A Primer of Real Functions*, 3rd edition, Carus Mathematical Monographs of the MAA, No. 13, 1981, pp. 141–142.
2. M. W. Botsko, An elementary proof that a bounded a.e. continuous function is Riemann integrable, *Amer. Math. Monthly* 95 (1988), 249–252.
3. R. R. Goldberg, *Methods of Real Analysis*, Blaisdell Publishing Co., New York, 1964, pp. 163–164.

Another Eulerian-Type Proof

DENNIS C. RUSSELL
York University
Toronto, Canada M3J 1P3

Euler's demonstration [1] that $\sum_{n=0}^{\infty}(2n+1)^{-2}=\pi^2/8$ is ingenious, though it does assume the value of $\int_0^1 x^{2n+1}(1-x^2)^{-1/2}\,dx$ and a power series for arcsin x, both of which require prior preparation. How about the following, where we use only Euler's formula $e^{ix}=\cos x+i\sin x$ and the well-known logarithmic series?

$$\begin{aligned}\int_0^{\pi/2}\log(2\cos x)\,dx &= \int_0^{\pi/2}\log\left[e^{ix}(1+e^{-2ix})\right]dx\\ &= \int_0^{\pi/2}\left[ix-\sum_{n=1}^{\infty}\frac{(-1)^n}{n}e^{-2nix}\right]dx\\ &= i\frac{\pi^2}{8}-\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\left[\frac{e^{-2nix}}{2ni}\right]_0^{\pi/2}\\ &= i\left[\frac{\pi^2}{8}-\sum_{m=0}^{\infty}\frac{1}{(2m+1)^2}\right].\end{aligned}$$

(The even terms disappear because $e^{-4mi(\pi/2)}-1=0$.) Since the left side is real and the right side purely imaginary, both of them are zero. Thus, as well as the required series, we get as a bonus that

$$\int_0^{\pi/2}\log(2\cos x)\,dx=0,\quad \text{namely}\quad \int_0^{\pi/2}\log\cos x\,dx=-\frac{\pi}{2}\log 2.$$

Another advantage of the above is that a few more words justify it completely, since the logarithmic series converges uniformly on any interval $[0, a]$, $0<a<\pi/2$, and we may let $a\to\pi/2-0$ by Abel's Continuity Theorem.

REFERENCE

1. Gerald Kimble, Euler's other proof, this MAGAZINE 60 (1987), 282.

PROBLEMS

LOREN C. LARSON, *editor*
St. Olaf College

GEORGE GILBERT, *associate editor*
Texas Christian University

Proposals

To be considered for publication, solutions should be received by May 1, 1992.

1383. *Proposed by Michael Handelsman, Erasmus Hall High School, Brooklyn, New York.*

Integers from 1 to n are randomly generated (as by a computer). Each integer has an equal probability of being selected, and unlimited repetition is permitted. As integers are generated, a running sun is recorded.

Given any integer k, such that $1 \leqslant k \leqslant n$, what is the probability that a sum of *exactly* k will be reached?

1384. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, Maryland.*

A square matrix M is *periodic* if and only if there is a positive integer r and a nonnegative integer s such that $\mathrm{M}^{r+s} = \mathrm{M}^{s}$. Characterize those fields K such that every square matrix over K is periodic.

1385. *Proposed by Howard Morris, Chatsworth, California.*

Let $f(x) = \prod_{n=1}^{\infty}(1 + x/2^n)$. Show that at the point $x = 1$, $f(x)$ and all its derivatives are irrational.

1386. *Proposed by Jiro Fukuta, Motosu-gun, Gifu-ken, Japan.*

Let ABC be an acute-angled triangle, let H be the foot of the altitude from A, and let D, E, Q be the feet of the perpendiculars from an arbitrary point P in the triangle onto AB, AC, AH, respectively. Prove that

$$|AB \cdot AD - AC \cdot AE| = BC \cdot PQ,$$

where $AB, AD, \dots$ denote the length of segments $AB, AD, \dots$.

ASSISTANT EDITORS: CLIFTON CORZAT, BRUCE HANSON, RICHARD KLEBER, KAY SMITH, and THEODORE VESSEY, *St. Olaf College* and MARK KRUSEMEYER, *Carleton College. We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) next to a problem number indicates that neither the proposer nor the editors supplied a solution.*

Solutions should be written in a style appropriate for Mathematics Magazine. Each solution should begin on a separate sheet containing the solver's name and full address.

Solutions and new proposals should be mailed in duplicate to Loren Larson, Department of Mathematics, St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057-1098 or mailed electronically via fax: (507) 663-3549 or e-mail: larson@stolaf.edu.

1387. *Proposed by Kenneth Stolarsky, University of Illinois at Urbana-Champaign, Urbana, Illinois.*

Given $\varepsilon > 0$ and a function $f(x)$, continuous on $(-\infty, \infty)$, must there exist a function $g(x)$, continuous on $[0,1]$, such that

$$\inf_{-\infty < y < \infty} \left(\max_{0 \leqslant x \leqslant 1} |g(x) - f(x+y)| \right) \geqslant \varepsilon?$$

Quickies

Answers to the Quickies are on page 357.

Q784. *Proposed by Michael Golomb, Purdue University, West Lafayette, Indiana.*

Let Δ be a triangle subdivided into triangles $\Delta_1, \Delta_2, \dots, \Delta_n$. Let $A, A_1, A_2, \dots, A_n$ denote the areas of the ellipses of maximum area inscribed in $\Delta, \Delta_1, \Delta_2, \dots, \Delta_n$, respectively. Show that

$$A = A_1 + A_2 + \cdots + A_n.$$

Q785. *Proposed by Jeffrey Shallit, University of Waterloo, Waterloo, Canada.*

Prove that every real number in the interval $[0,2]$ can be written as the sum of two elements of the Cantor set.

Solutions

Generalized Pentagonal Numbers **December 1990**

1358. *Proposed by Peter Ross, Santa Clara University, Santa Clara, California.*

Which positive integers are representable in the form $\binom{k}{2} + kn$, $k > 1$, $n \geqslant 1$?

Solution by Jack V. Wales, Jr., The Thacher School, Ojai, California.

A positive integer x is representable in the given form if and only if x is not a power of 2, that is, if and only if x has an odd factor greater than 2. Further, any such x has as many such representations as there are distinct factorizations of $2x$ into two distinct factors, one of which is odd.

Let x be a number of the form $\binom{k}{2} + nk$, $k > 1$, $n \geqslant 1$. Then $x = k(k-1)/2 + nk$, or

$$2x = k(k + 2n - 1).$$

We see that $2x$ must have an odd factor greater than 2, and therefore so must x.

Now suppose x is any number with an odd factor greater than 2. Consider any factorization of $2x$ into two distinct factors, one of which is odd. Let A be the smaller of the two factors and B the larger. Then $k = A$ and $n = (B + 1 - A)/2$ will allow

$$\binom{k}{2} + kn = \frac{A(A-1)}{2} + \frac{A(B+1-A)}{2} = \frac{AB}{2} = x.$$

Also solved by Christos Athanasiadis (student), Seung-Jin Bang (Korea), Pierre Barnouin (France), Martin Bazant (student), Brian D. Beasley, Karl A. Beres, Ada Booth, W. E. Briggs, Stephen Bronn, David Callan, Con Amore Problem Group (Denmark), Steven W. Davis, Robert Doucette, Donald Duncan, Roger B. Eggleton (Brunei Darussalam), Milton P. Eisner, Kevin Ford (student), Arthur H. Foss, Robert Gallant (student, Canada), E. C. Greenspan and S. A. Greenspan, H. Guggenheimer, Robert Heller, Russell Jay Hendel, Francis M. Henderson, R. High, R. Daniel Hurwitz, Douglas E. Iannucci, Julie Kerr (student), Emil F. Knapp, David W. Koster, H. K. Krishnapriyan, Lamar University Problem Solving Group, Adam H. Lewenberg, Carl Libis, Graham Lord, Jack McCown, Nick MacKinnon (England), Helen M. Marston, James Martino, Thomas E. Moore, Roger B. Nelsen, C. C. Oursler, M. Parmenter (Canada), Stephen G. Penrice, Glenn Powers, Ken Rebman, Thomas Rike, Ioan Sadoveanu, Ramchander Sastry and Krishna Subramanian, Harvey Schmidt, Jr., Heinz-Jürgen Seiffert (Germany), John S. Sumner, Gan Wee Teck (student, Singapore), Ravi Vakil (student, Canada), David C. Vella, Jesús Selva Vera (Spain), Edward T. H. Wang and Wan-Di Wei, Richard Weida, Westmont College Problem Solving Group, Whittier College Problem Solving Group, and the proposer.

Ramchander Sastry and Krishna Subramanian and the *Con Amore Problem Group* pointed out that

$$\binom{k}{2} + kn = \binom{n+k}{2} - \binom{n}{2} = n + (n+1) + \cdots + (n+k-1),$$

so the problem is equivalent to determining the natural numbers expressible as a sum of consecutive integers, with more than just one term. This problem was posed by Sylvester, and is a problem in Section 1.3 of Niven and Zuckerman's *Introduction to the Theory of Numbers*, and has a similar easy solution described by Melfried Olson in "Sequentially So," this MAGAZINE, 52(1979), 297–298. *Callan* notes that the problem is equivalent to a problem involving a sequentially defined sequence that appears in the *American Mathematical Monthly* (E3266, May 1988, Solution, October 1989).

Diagonal inequality in a positive definite matrix **December 1990**

1359. *Proposed by Frank Schmidt, Bryn Mawr College, Bryn Mawr, Pennsylvania.*

Let $\mathbf{A} = (a_{ij})$ be a real $n \times n$ positive definite matrix.

(a) Show that $\prod_{i=1}^{n} a_{ii} > \prod_{i=1}^{n} a_{i\sigma(i)}$ for all $\sigma \in S_n$, $\sigma \neq$ identity.

(b) Show that $\sum_{i=1}^{n} a_{ii} > \sum_{i=1}^{n} a_{i\sigma(i)}$ for all $\sigma \in S_n$, $\sigma \neq$ identity.

Solved by Ioan Sadoveanu, Ellensburg, Washington.

We will show that if $\mathbf{A} = (a_{ij})$ is a complex $n \times n$ positive definite matrix then

$$\prod_{i=1}^{n} a_{ii} > \prod_{i=1}^{n} |a_{i\sigma(i)}| \quad \text{and} \quad \sum_{i=1}^{n} a_{ii} > \sum_{i=1}^{n} |a_{i\sigma(i)}|$$

for all $\sigma \in S_n$, $\sigma \neq$ identity.

Since all principal minors of a positive definite matrix are positive numbers,

$$\det(a_{ii}) > 0,\ i = 1,2,\ldots,n \quad \text{and} \quad \det\begin{pmatrix} a_{ii} & a_{i\sigma(i)} \\ a_{\sigma(i)i} & a_{\sigma(i)\sigma(i)} \end{pmatrix} > 0, \quad \sigma(i) \neq i.$$

Hence

$$a_{ii} > 0,\ i = 1,2,\ldots,n \quad \text{and} \quad a_{ii}a_{\sigma(i)\sigma(i)} > a_{i\sigma(i)}a_{\sigma(i)i} = a_{i\sigma(i)}\overline{a_{i\sigma(i)}} = |a_{i\sigma(i)}|^2 \quad \text{for} \quad \sigma(i) \neq i.$$

Thus,

$$\prod_{i=1}^{n} a_{ii} = \sqrt{\prod_{i=1}^{n} a_{ii} a_{\sigma(i)\sigma(i)}} > \sqrt{\prod_{i=1}^{n} |a_{i\sigma(i)}|^2} = \prod_{i=1}^{n} |a_{i\sigma(i)}|,$$

and

$$\sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \frac{a_{ii} + a_{\sigma(i)\sigma(i)}}{2} \geqslant \sum_{i=1}^{n} \sqrt{a_{ii} a_{\sigma(i)\sigma(i)}} > \sum_{i=1}^{n} \sqrt{|a_{i\sigma(i)}|^2} = \sum_{i=1}^{n} |a_{i\sigma(i)}|.$$

When **A** is a real positive definite matrix, the result of the problem follows, because $|x| \geqslant x$ for all real x.

Also solved by Christos Athanasiadis (student), David Callan, Con Amore Problem Group (Denmark), Murray S. Klamkin, Julio Kuplinsky, Marvin Marcus, Heinz-Jürgen Seiffert (Germany), John S. Sumner, Zhongcheng Wang (student) and Weimin Xue (China), The Xavier Mathematics Problem League, and the proposer.

The proposer is interested in references or information concerning the following open question: Given n real vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$, (each $\mathbf{b}_i \in \mathbf{R}^k$), which $\sigma \in S_n$ gives the *smallest* inner product $\Sigma_{i=1}^k \mathbf{b}_i \cdot \mathbf{b}_{\sigma(i)}$?

Characterization of a disc **December 1990**

1360. *Proposed by Paul R. Scott, University of Adelaide, Adelaide, Australia.*

Let A be a bounded convex set in the plane and B a set congruent to it. Given $A \cap B$ is a centrally symmetric set for all positions of A and B, must A be a circular disc?

I. *Solution by John S. Sumner, University of Tampa, Tampa, Florida.*

Not necessarily. If A is a line segment in the plane, then A is trivially bounded and convex. If B is a line segment congruent to A, then for any positions of A and B, the set $A \cap B$ is either empty, a point, or another line segment. In either case $A \cap B$ is centrally symmetric.

II. *Solution by R. High, New York City, New York.*

The answer is yes, provided we add the requirement that A has a nonempty interior (e.g., three non-collinear points).

Note that every point x on the boundary of A has a well-defined left and right tangent, L_x and R_x respectively, the limits of the family of all lines through x not meeting A in any other point.

We first establish that A cannot have a line segment as part of its boundary. Suppose it does. Since the boundary of A cannot be completely straight, consider a portion that is not straight. For small enough $\varepsilon > 0$, we can choose two "chords" CD and $C'D'$, each of length ε, CD along the straight portion of the boundary of A and $C'D'$ along the non-straight portion (coincident with the boundary only at the endpoints C' and D'). Now choose a set B congruent to A and align B on A so that the chords CD and $C'D'$ coincide. It is clear that in this position the intersection $A \cap B$ is not centrally symmetric.

We next show that for sufficiently small $\varepsilon > 0$, all "arcs" of the boundary of A with chords of length ε are congruent. For suppose CD and $C'D'$ are chords of A. Choose B congruent to A and align B so that the two chords CD and $C'D'$ coincide in such a way that $A \cap B$ is the union of two "slivers" joined at the chord. Some thought shows that the center of symmetry of $A \cap B$ must be at the midpoint of this common chord (this chord is necessarily the longest segment of $A \cap B$ parallel to the support lines of $A \cap B$ in this direction).

We claim that the boundary of A cannot have any "corners" (points where L_x is discontinuous). For suppose x were a corner of A, where L_x jumps by, say δ, as x traverses the boundary of A. Then we could find an $\varepsilon > 0$ and a chord of length ε such that the corresponding arc contains only corners smaller than δ (since there are only finitely many as large as δ). But the two arcs could not then be congruent.

Now choose a small $\varepsilon > 0$ such that a sequence of chords of length ε define an equilateral polygon inscribed in A. This polygon must be *equiangular*. For if not, consider the behavior of L_x as it approaches two vertices with distinct angles. Since all arcs are congruent, one of the two vertices would have to be a corner of A, which we've seen cannot happen.

Finally, note that the only curve approximated by inscribed regular polygons of an arbitrary number of sides is the circle.

Also solved by Ioan Sadoveanu, the Xavier Mathematics Problem League, and the proposer.

Composition of differentiable functions **December 1990**

1361. *Proposed by Mark Krusemeyer, Carleton College, Northfield, Minnesota.*

Does there exist a differentiable function f, defined for all real numbers x, such that $f(f(x)) = e^x$ for all x? If so, exhibit such a function; if not, show why not.

I. *Solution by Reiner Martin, student, University of California, Los Angeles, California.*

Yes, such a function exists.

Let $\exp^{(n)}(x)$ be defined, for integers n, by $\exp^{(0)}(x) = x$ and $\exp^{(n+1)}(x) = \exp(\exp^{(n)}(x))$. For example, we have $\exp^{(-2)}(x) = \log\log x$. Then it is straightforward to check that

$$f(x) = \begin{cases} \exp^{(n)}(\exp^{(-n)}(x) + 1/2) & \text{if } \exp^{(n)}(0) < x \leqslant \exp^{(n)}(1/2), \\ \exp^{(n+1)}(\exp^{(-n)}(x) - 1/2) & \text{if } \exp^{(n)}(1/2) < x \leqslant \exp^{(n+1)}(0) \end{cases}$$

defines a differentiable function, and that $f(f(x)) = \exp(x)$ for all real numbers x.

II. *Solution by David Callan, University of Wisconsin, Madison, Wisconsin*

Many such solutions exist. To investigate the possibilities for f, first note that since $f \circ f = \exp$ is one-to-one, f is also one-to-one and, thus, by continuity, f is (strictly) monotone.

LEMMA. *If (x, y) is on the graph of f, then so are (y, e^x) and (provided $y > 0$) $(\ln y, x)$.*

Proof of Lemma. $f(y) = f(f(x)) = e^x$; since $f(f)(\ln y)) = e^{\ln y} = y = f(x)$ and f is one-to-one, $f(\ln y) = x$.

Now let $a = f(0)$. If $a \notin (0,1)$, we can use the lemma to find points on the graph of f that violate monotonicity. Thus, $a \in (0,1)$ and $f(a) = f(f(0)) = 1$. Now if we join the points $(0, a)$ and $(a, 1)$ in any continuous increasing fashion to specify f, then the lemma serves to extend f (continuously) to the interval $[a, 1]$, then to $(1, e^a]$, $(e^a, e^1]$, $(e^1, e^{e^a}]$, and so on. Thus we have an increasing sequence (a_n) given by

$$a_0 = 0, \qquad a_{2k} = \exp^{(k)}(0), \qquad a_1 = a, \qquad a_{2k+1} = \exp^{(k)}(a)$$

(where superscripts denote iteration) such that $f\colon [a_i, a_{i+1}] \to [a_{i+1}, a_{i+2}]$. Working backwards using the inverse, ln, of exp, we also find f on $[\ln a, 0)$ and $(-\infty, \ln a)$, at which point the process stops.

To insure differentiability, start by choosing f to be smooth on the initial interval $[0, a]$, with $f' > 0$. Then $y = f(x)$ has a differentiable inverse $x = g(y)$ for $x \in (0, a)$. Hence for $y \in (a, 1)$, $f(y) = e^{g(y)}$, so f is also differentiable on $(a, 1)$, with $f'(y) = f(y)/(f'(g(y))$. By induction, we find that f is differentiable on the interior (a_i, a_{i+1}) of each interval (include the negative intervals by setting $a_{-2} = -\infty$, $a_{-1} = \ln a$), with

$$f'(y) = \frac{f(y)}{f'(f^{-1}(y))} \quad \text{for } y > a, \qquad f'(y) = \frac{e^y}{f'(f(y))} \quad \text{for } y < 0.$$

Finally, we must consider the endpoints of intervals. Since $f'(a^+) = 1/f(0^+)$, we can insure the differentiability of f at a by choosing f to satisfy $f'(a^-) = 1/f'(0^+)$. By induction, f will then also be differentiable at the other endpoints, since $f'(a_{n+1}^+) = a_{n+2}/f'(a_n^+)$ and $f'(a_{n+1}^-) = a_{n+2}/f'(a_n^-)$.

Also solved by Irl C. Bivens and Benjamin G. Klein, Duane Broline, P. A. de Caux, D. K. Cohoon, Con Amore Problem Group (Denmark), David Goering, Stephen Noltie, John S. Sumner and Kevin L. Dove, Western Maryland College Problems Group, and the proposer. Two incorrect solutions were received.

In addition to giving a solution, *Broline* communicated the information from *Hillel Gauchman* that the problem is known. References include Kuczma's book, *Functional Equations in a Single Variable* (PWN, Warsaw, 1968), and papers by *H. Kneser* (*J. reine u. angew Math.* 187 (1950), 56–67) and others. *Murray Klamkin* noted that the result is contained in the solution given by *I. N. Baker* to a problem of *D. J. Newman's* in the *American Mathematical Monthly* 62 (1955), 190–191.

Bivens and Klein, Broline, de Caux, Cohoon, Goering, Sumner and Dove, and the *WMC Problems Group* all showed that infinitely many f exist. The solutions by *de Caux, Noltie, Sumner and Dove,* and the proposer, all included the specific f from *Martin's* solution given above. *Cohoon* and the *WMC Problems Group* considered more general functional equations of the form $f(f(x)) = g(x)$.

Inequality for n triangles **December 1990**

1362. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

If (a_i, b_i, c_i) are sides, R_i the circumradii, r_i the inradii, and s_i the semi-perimeters of n triangles $(i = 1, 2, \dots, n)$ respectively, show that

$$3\left(\prod a_i^{-1/n} + \prod b_i^{-1/n} + \prod c_i^{-1/n}\right) \leqslant \prod \left(\frac{s_i}{r_i R_i}\right)^{1/n}$$
$$\leqslant 2^n \prod \left(a_i^{-1/n} + b_i^{-1/n} + c_i^{-1/n}\right),$$

where the sums and products are over $i = 1$ to n.

(We regret that the statement of this problem inadvertantly contained an extra summation sign, which rendered the problem meaningless, and omitted the necessary condition that $n > 1$.)

Solution by Jiro Fukuta, Motosu-gun, Gifu-ken, Japan.

From Hölders inequality,

$$3\left(\prod a_i^{-1/n} + \prod b_i^{-1/n} + \prod c_i^{-1/n}\right) \leqslant 3 \prod \left(a_i^{-1} + b_i^{-1} + c_i^{-1}\right)^{1/n}.$$

Now $(a_i + b_i + c_i)^2 = a_i^2 + b_i^2 + c_i^2 + 2(a_i b_i + b_i c_i + c_i a_i) \geqslant 3(a_i b_i + b_i c_i + c_i a_i)$ and therefore $3(a_i^{-1} + b_i^{-1} + c_i^{-1}) \leqslant (a_i + b_i + c_i)^2/(a_i b_i c_i) = s_i/(r_i R_i)$. It follows that

$$3\left(\prod a_i^{-1/n} + \prod b_i^{-1/n} + \prod c_i^{-1/n}\right) \leqslant \prod \left(\frac{s_i}{r_i R_i}\right)^{1/n}.$$

For the other inequality, it is enough to prove that if a, b, c are the lengths of the three sides of a triangle,

$$\left(\frac{(a+b+c)^2}{abc}\right)^{1/n} \leqslant 2(a^{-1/n}+b^{-1/n}+c^{-1/n}),$$

or

$$(a+b+c)^2 \leqslant \left(2(bc)^{1/n}+2(ca)^{1/n}+2(ab)^{1/n}\right)^n.$$

Let $A = a^{1/n}$, $B = b^{1/n}$, and $C = c^{1/n}$. Then the preceding equation can be rewritten as

$$(A^n+B^n+C^n)^2 \leqslant (2BC+2CA+2AB)^n.$$

But

$$\begin{aligned}(2BC+2CA+2AB)^n &= \left(A(B+C-A)+B(C+A-B)+C(A+B-C)+A^2+B^2+C^2\right)^n \\ &\geqslant (A^2+B^2+C^2)^n\end{aligned}$$

because $(B+C)^n = (b^{1/n}+c^{1/n})^n \geqslant b+c \geqslant a = A^n$, and thus $B+C \geqslant A$, and so on. Applying Jensen's inequality, we have

$$(A^n+B^n+C^n)^2 \leqslant (A^2+B^2+C^2)^n,$$

for $n > 0$. The result follows.

Also solved by Ioan Sadoveanu and the proposer.

Comments

Q774. In the Quickie Solution (Vol. 64, No. 1, p. 61, 67) to this problem, the author states that "$\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are in consecutive order on the circle and do not lie on any semicircle." The following example, pointed out by *Mike Schramm (student)* and *Kevin Farrell, Lyndon State College*, shows that this conclusion is false. Let $\mathbf{A} = (5, 14)$, $\mathbf{B} = (14, -5)$, $\mathbf{C} = (-5, 14)$, and $\mathbf{D} = (14, 5)$. These vectors satisfy the hypotheses of the theorem. namely, they are distinct coplanar vectors with equal lengths and $\mathbf{A}\cdot\mathbf{B}+\mathbf{C}\cdot\mathbf{D} = \mathbf{A}\cdot\mathbf{D}+\mathbf{B}\cdot\mathbf{C}$. They do not satisfy the conclusion stated above. They are not in consecutive order and do lie on a semicircle.

Murray Klamkin, University of Alberta comments: In the solution, it was assumed that AC and BD, which must be perpendicular chords of a circle, intersect at a point P lying within or on a circle. The example given above shows that P might lie outside the circle. Here is a simpler solution that takes care of both cases.

Choose a rectangular coordinate system with origin at the center O of the circle and with axes parallel to AC and BD. Then the four vectors have the representations: $\mathbf{A} = (\alpha, \beta)$, $\mathbf{C} = (-\alpha, \beta)$, $\mathbf{B} = (\gamma, \delta)$, and $\mathbf{D} = (\gamma, -\delta)$. (Note that α need not be positive, etc.) Finally,

$$\mathbf{A}\cdot\mathbf{B}+\mathbf{C}\cdot\mathbf{D} = \alpha\gamma+\beta\delta-\alpha\gamma-\beta\delta = 0.$$

Since also

$$\mathbf{A}^2+\mathbf{B}^2+\mathbf{C}^2+\mathbf{D}^2-(\mathbf{A}-\mathbf{B})^2-(\mathbf{C}-\mathbf{D})^2 = 2(\mathbf{A}\cdot\mathbf{B}+\mathbf{C}\cdot\mathbf{D}) = 0,$$

we have equivalently that

$$AP^2 + BP^2 + CP^2 + DP^2 = 4R^2, \quad (R = \text{radius of the circle}).$$

The latter corresponds to the known result (*Crux Mathematicorum* 15 (1989) 293, #1) that the sum of the areas of the four circles whose diameters are AP, BP, CP, and DP is equal to the area of the given circle. In this result it is assumed that P lies within the circle. But the above proof shows that it is valid if P is outside the circle. This four-circle result apparently has been generalized (*Crux Mathematicorum* 16 (1990), p. 109, #1535) to a result concerning two intersecting chords in an ellipse. However, the ellipse result can be shown to follow from the circle result by an affine transformation.

Answers

Solutions to the Quickies on page 351.

A784. By an affine transformation, one can transform a given triangle T into an equilateral triangle. Such a transformation transforms inscribed ellipses into inscribed ellipses and preserves ratios of areas. Therefore, if $A(T)$ is the area of T and $A(E)$ is the maximal area of an inscribed ellipse,

$$A(E) = \gamma A(T),$$

for some constant γ independent of T. The claim follows. (For the equilateral triangle, the inscribed ellipse of maximal area is the inscribed circle, hence $\gamma = \pi\sqrt{3}/9$.)

A785. Recall that the Cantor set C is the set of all real numbers in $[0,1]$ containing only the digits 0 and 2 in their base-3 representation. Thus we can state the result as $[0,2] \subseteq C + C$.

Let D be the set of all real numbers containing only the digits 0 and 1 in their base-3 representation. It is easy to see that $[0,1] \subseteq D + D$. For example, if $x \in [0,1]$ then we write its base-3 representation as

$$x = .x_1x_2x_3x_4 \dots .$$

Then we put

$$y_i = \begin{cases} 1, & \text{if } x_i \geqslant 1, \\ 0 & \text{otherwise}. \end{cases}$$

$$z_i = \begin{cases} 1, & \text{if } x_i \geqslant 2, \\ 0 & \text{otherwise}. \end{cases}$$

Then $x_i = y_i + z_i$, so $x = y + z$ (with no carries!).

To prove $[0,2] \subseteq C + C$, we take an $x \in [0,2]$. Express $x/2 = y + z$, with $y, z \in D$. Then $x = 2y + 2z$. But $2y$ and $2z$ are numbers containing only the digits 0 and 2 in their base-3 representation. (Note: Since $C + C \subseteq [0,2]$, it follows that $C + C = [0,2]$.)

This result is not new (but the proof seems to be). It forms the basis of *Marshall Hall's* result in his paper "On the sum and product of continued fractions" (*Ann. Math.* (2) 48 (1947), 966–993). Also, it is a theorem of *E. Borel* (*Éléments de la Théorie des Ensembles, Paris*, 1949, Note V). Also, see *M. Pavone*, The Cantor set and a geometric construction (*L'enseignement Math.* 35 (1989), 41–49).

REVIEWS

PAUL J. CAMPBELL, *editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of the mathematics literature. Readers are invited to suggest items for review to the editors.

Maran, Stephen P., Science is dandy, but promotion can be lucrative, *Smithsonian* 22(2) (May 1991) 72–80.

"Making a discovery is only half the battle; announcing it to one's advantage is the next step for the scientifically ambitious." This article explores the tension between the professional obligation to reveal scientific discoveries promptly and the desire of scientists to exploit privately their own discoveries—and hence derive greater personal benefit. But the benefits accrue only to those who know how to manage the publicity. Among the lessons is one close to home for mathematicians: J.C. Adams and U.J.J. Le Verrier deduced by mathematical calculation the existence of Neptune. "But they lacked the personal printing press of Brahe, the self-promotion skills of Galileo and the intercession of the scientific establishment that Herschel enjoyed"—so Adams and Le Verrier were simply ignored, until chance intervened.

Kanigel, Robert, *The Man Who Knew Infinity: A Life of the Genius Ramanujan*, Scribners, 1991; ix + 438 pp, $27.95. ISBN 0-684-19259-4

An *absolutely splendid* and inspired book, by a sensitive and talented biographer, perfectly suited to be read by mathematicians and non-mathematicians alike.

Bigger slices of pi, *Science News* 140 (24 August 1991) 127.

New record for pi: 2.16 billion decimal digits, by David V. and Gregory V. Chudnovsky (Columbia University).

Maurer, Stephen B., and Anthony Ralston, *Discrete Algorithmic Mathematics*, Addison-Wesley, 1991; xix + 898 pp. ISBN 0-201-15885-0

Tony Ralston stimulated a great river of curriculum change—including today's backwash (mostly talk) of calculus "reform"—by questioning the primacy of calculus and urging the inclusion of discrete mathematics in the canon of the first two years of college mathematics. Steve Maurer, with the backing of the Sloan Foundation, promoted discrete mathematics to the mathematical community. Here is their vision of what should be done in a one-year discrete freshman or sophomore mathematics course, intended for the same students who are well-prepared for calculus. The central theme is algorithms, and the central methods are induction and recursion. After chapters on algorithms, induction, graphs and trees, combinatorics, difference equations, probability, logic, algorithmic linear algebra, and infinite processes, the book culminates in a comparison of sorting algorithms.

Michaels, John G., and Kenneth H. Rosen, *Applications of Discrete Mathematics*, McGraw-Hill, 1991; x + 516 pp, (P). ISBN 0-07-041823-3

Twenty-four chapters cover "a wide variety of interesting applications of discrete mathematics," intended as a supplement for students studying discrete mathematics or as a text for a second course in the subject (particularly for a summer course for high-school teachers). Each chapter includes exercises classified by difficulty (with solutions), a bibliography for further reading, the historical background of the application, and a set of computer projects. Readers of the book are invited to suggest topics for additional chapters on further applications, and even to submit such chapters for possible inclusion in future editions.

Tierney, John, Behind Monty Hall's doors: puzzle, debate, and answer?, *New York Times* (21 July 1991) 1, 9. Gillman, Leonard. 1991. The car-and-goats fiasco. *Focus* (June-July 1991): 8.

"Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the other doors, opens another door, say No. 3, which has a goat. He then says to you, 'Do you want to pick door No. 2?' Is it to your advantage to take the switch?" Marilyn vos Savant described this puzzle in *Parade* in September, 1990; she received thousands of vitriolic letters—many from mathematicians—objecting to her counterintuitive answer of "yes." Leonard Gillman carefully analyzes several variations of the game and castigates those mathematicians not only for ignorance but also for arrogance and condescension. (Perhaps it shouldn't be possible to get a Ph.D. in mathematics without taking a course in probability and one in humility.) Think of the damage this problem has been doing to the credibility and image of mathematicians! Monty Hall, host of the TV show "Let's Make a Deal," conducted a simulation with a reporter as contestant: 10 rounds of non-switching produced 4 cars, 6 goats; 10 rounds of switching produced 8 cars, 2 goats. But 10 rounds of Monty Hall "controlling" the game yielded 0 cars: Whenever the reporter first picked a door with a goat, Hall opened it and gave him the goat—no chance to switch; when the reporter picked the door with the car, Hall would try to bribe him to switch doors. Hall used a loophole that some read into the problem statement, that the host need not open a door or offer the switch. Hall's advice: "[I]f you can get me to offer you $5,000 not to open the door, take the money and go home."

Mirollo, Renato, and Steven Strogatz, Synchronization of pulse-coupled biological oscillators, *SIAM Journal on Applied Mathematics* 50(6) (December 1990) 1645–1662. Stewart, Ian, All together now ..., *Nature* 350 (18 April 1991) 557. Peterson, Ivars, Step in time: exploring the mathematics of synchronously flashing fireflies, *Science News* 140 (31 August 1991) 136–137.

In Southeast Asia there are huge swarms of fireflies who flash *in unison*. Why? How? There are other examples of biological synchrony: "the pacemaker cells of the heart, networks of neurons in the circadian pacemaker and hippocampus, the insulin-secreting cells of the pancreas, crickets that chirp in unison, and groups of women whose menstrual cycles become mutually synchronized." But what about the fireflies? Mirollo and Strogatz have proved a conjecture of C.S. Peskin that "synchrony is the rule if every firefly interacts with every other." The fireflies, as oscillators, are mutually coupled through seeing each other's signals; eventually all the oscillations become synchronized. In a more general network, the fireflies synchronize even if some can't see some of the others, provided there is no isolated group that cannot see any of the rest. (Ah! New mathematics that you can bring up at your next cocktail party!)

Freedman, David H., A chaotic cat takes a swipe at quantum mechanics, *Science* 253 (9 August 1991) 626.

Although quantum mechanics is completely deterministic, it is supposed to agree with classical physics on the macro level, where chaos can happen. Joseph Ford (Georgia Tech), a chaos theorist, experimented the set of equations called Arnol'd's cat, which exhibits chaotic behavior at the macro level. After shrinking down to the micro level and quantizing the system, he restored it to the macro level by gradually decreasing Planck's constant. The reconstituted system, contrary to expectation, does not exhibit any chaotic behavior, though it should. Ford expects to support his challenge to quantum mechanics with an experiment, while believers in quantum mechanics expect to find fault with his proof.

Cipra, Barry A., A fractal focus on the Mandelbrot set, *SIAM News* 24(4) (July 1991) 22.

Mitsuhiro Shishikura (SUNY—Stony Brook) has proved that the boundary of the Mandelbrot set has Hausdorff dimension 2. So do Julia sets associated with boundary points, though some have (two-dimensional) Lebesgue measure 0. Still unproven: that the boundary of the Mandelbrot set has Lebesgue measure 0, or that it is locally connected.

Cipra, Barry, A speedier way to decompose polygons, *Science* 253 (19 July 1991) 270.

Decomposing polygons is a "key procedure in computational geometry," for such purposes as the finite-element method and the painting of surfaces in CAD/CAM solid-modeling systems. Bernard Chazelle (Princeton University) has devised an $\mathcal{O}(N)$ algorithm for the problem, besting $\mathcal{O}(N^2)$ for naive algorithms and the previous record of $\mathcal{O}(N \log N)$.

Solow, Daniel, *How to Read and Do Proofs: An Introduction to Mathematical Thought Processes*, 2nd ed., Wiley, 1990; xx + 242 pp, \$21.95 (P). ISBN 0-471-51004-1

Second edition of the most important book for a mathematics major to read and master. A new chapter has been added on nested quantifiers, and the number of exercises has been doubled. All examples and exercises are drawn from high-school mathematics.

Glimm, James G., *Mathematical Sciences, Technology, and Economic Competitiveness*, National Academy Press, 1991; x + 114 pp \$22 (US), \$26.50 (export) (P). ISBN 0-309-04483-9

Analyzes in detail the role of mathematics in five key American industries (aircraft, semiconductors and computers, petroleum, cars, and telecommunications). Stresses the key roles of the mathematical sciences in all stages of the product cycle and notes specific recent mathematical discoveries that have contributed to the "technology base."

Peterson, Ivars, The checkers challenge: A checker-playing computer program contends for the world title, *Science News* 140 (3) (20 July 1991) 40–41.

Checkers hasn't the glamor of its more cerebral cousin chess and so has suffered comparative neglect by programmers. Now there is Chinook, by Jonathan Schaeffer (University of Edmonton), which usually draws against Marion Tinsley, a semiretired mathematician who has been world checkers champion whenever he has cared to (1956–58 (resigned), 1975–91). When the American and British checkers organizations refused to sanction a title match between Chinook and Tinsley, Tinsley—who has lost only *three* of the thousands of games he has played since 1975—resigned his title again, in protest.

NEWS AND LETTERS

LETTERS TO THE EDITOR

Dear Editor:

Readers of this MAGAZINE may have noticed that Ayoub B. Ayoub's "Proof without Words: The Reflection Property of the Parabola" (June 1991) is very old. It was known to the ancient Greeks. They didn't use coordinate systems or calculus, but why should one? The parabola is the set of all points that have equal distances to a fixed point (the focus) and a straight line, a fact used in Ayoub's proof. Then $m_1 \cdot m_2 = -1$ (orthogonality) is a geometric triviality known since ancient times. The conclusion is the same, but what were coordinate systems and derivatives used for?

I fight against the deplorable fact that many of our students today know parabolas only as graphs of square or square root functions. Many students who will not go into mathematics do not understand the subtle notions of calculus, they just apply the machinery. The solution to the tangent problem for the conic sections can be understood by anyone with minimal knowledge of geometry. And it teaches to understand what a tangent is, something many students do not fully appreciate.

Dr. Andreas Müller
Zürich, Switzerland

Dear Editor:

Readers who enjoyed James Fife's article "The Kuratowski Closure-Complement Problem" (June 1991) might also enjoy Elementary Problem E3144 in the *American Mathematical Monthly* (stated Vol. 93, No. 4, April 1986; solution Vol. 95, No. 4, April 1988). This problem is similar to the Kuratowski problem, but with the operations of taking the complement, interior, and boundary of a set. At most, 34 sets can be so constructed.

Another similar problem in the *Monthly* is Advanced Problem 5996 (stated Vol. 81, No. 9, November 1974; solution Vol. 85, No. 4, April 1978). This problem asks the same question for the operations closure, interior, and union, and results in 13 sets. Eric Langford's article "Characterization of Kuratowski 14-Sets" (*Monthly*, Vol. 78, No. 4, April 1971) might also be of interest.

John P. Robertson
Berwyn, PA

ACKNOWLEDGEMENTS

In addition to our associate editors, the following have assisted the MAGAZINE *by refereeing papers during the past year. We appreciate the time and care they have given.*

Alavi, Yousef, *Western Michigan University*
Alberts, George E., *National Security Agency, Fort George Meade, MD*
Alexanderson, Gerald L., *Santa Clara University*
Aminzadeh, Mostafa, *Towson State University*
Andrews, George E., *Pennsylvania State University, University Park*
Ashford, Stella Roberson, *Southern University*
Askey, Richard, *University of Wisconsin, Madison*
Banchoff, Thomas, *Brown University*
Barbeau, Edward J., *University of Toronto*
Barksdale, James B., Jr., *Western Kentucky University*
Barnette, Dave, *University of California, Davis*
Bauman, Steven, *University of Wisconsin, Madison*
Baxley, John V., *Wake Forest University*
Beals, Richard W., *Yale University*
Beaver, Olga R., *Williams College*
Beezer, Robert A., *University of Puget Sound*
Berndt, Bruce C., *University of Illinois*
Bloch, Norman, *SUNY College at Brockport*
Boas, Mary L., *DePaul University*
Boas, Ralph P., *Northwestern University*
Borrelli, Robert, *Harvey Mudd College*
Borwein, Jonathan M., *Dalhousie University*
Braden, Bart, *Northern Kentucky University*
Branton, Richard G., *West Chester University*
Brown, Jack B., *Auburn University*
Brualdi, Richard A., *University of Wisconsin, Madison*
Brunson, Barry W., *Western Kentucky University*

Bushaw, Donald W., *Washington State University*
Cammack, Larry, *Central Missouri State University*
Carlip, Walter, *University of Chicago*
Castellanos, Dario, *University of Carabobo*
Chae, Soo Bong, *New College*
Chakerian, G.D., *University of California, Davis*
Chen, Henry, *Towson State University*
Chollet, John, *Towson State University*
Connelly, Robert, *Cornell University*
Coughlin, Jim, *Towson State University*
Cunningham, Stephen, *California State University, Stanislaus*
Dancis, Jerry, *University of Maryland, College Park*
Dauben, Joseph W., *Herbert H. Lehman College*
Delany, Jim, *California Polytech State University*
Dobbs, David E., *University of Tennessee, Knoxville*
Drobot, Vlad, *San Jose State University*
Drucker, Thomas, *Dickinson College*
Farris, Frank, *Santa Clara University*
Fasanelli, Florence D., *Sidwell Friends School*
Fine, Benjamin, *Fairfield University*
Fisher, J. Chris, *University of Regina*
Foster, Lorraine L., *California State University, Northridge*
Fraser, Craig, *Victoria College*
Gallian, Joseph A., *University of Minnesota-Duluth*
Gardner, Martin, *Hendersonville, NC*
Ghahramani, Saeed, *Towson State University*
Gillman, Leonard, *University of Texas, Austin*
Gold, Bonnie, *Wabash College*
Goldstein, Larry J., *Goldstein Software, Brookeville, MD*
Goodson, Geoffrey, *Towson State University*
Graham, Ronald, *AT&T Bell Labs, Murray Hill, NJ*
Grant, John, *Towson State University*
Grassl, Richard M., *University of Northern Colorado*
Green, Euline, *Abilene Christian University*
Green, Judy, *Marymount University*
Grimaldi, Ralph, *Rose-Hulman Institute of Technology*
Grünbaum, Branko, *University of Washington*
Gulick, Denny, *University of Maryland, College Park*
Guy, Richard, *University of Calgary*
Halmos, Paul R., *Santa Clara University*
Hanson, Robert, *Towson State University*
Hatcher, Rhonda, *Texas Christian University*
Hawkins, Thomas W., *Boston University*
Heinrich, Katherine, *Simon Fraser University*
Herman, Eugene, *Grinnell College*
Hilton, Peter, *SUNY at Binghamton*
Hochwald, Scott, *University of North Florida*
Holden, Herb, *Gonzaga University*
Honsberger, Ross, *University of Waterloo*
Horak, Martin, *Towson State University*
Horn, Roger A., *The Johns Hopkins University*
Howard, F.T., *Wake Forest University*
Ilgenfritz, Ernie, *Towson State University*
Jacobson, Bernard, *Franklin & Marshall College*
Jepsen, Charles H., *Grinnell College*
Jesseph, Douglas, *Illinois Institute of Technology*
Job, Vanessa, *Marymount University*
Johnston, Elgin, *Iowa State University*
Jones, Sharon, *Towson State University*
Jordan, Jim, *Washington State University*
Karr, Alan F., *The Johns Hopkins University*
Kastner, Bernice, *Towson State University*
Kelly, Alice, *Santa Clara University*
Kemperman, J.H.B., *Rutgers University*
Kiltinen, John O., *Northern Michigan University*
Kim, Ohoe, *Towson State University*
Kimberling, Clark H., *University of Evansville*
Konhauser, Joseph D.E., *Macalester College*
Klonski, Leonard, *Santa Clara University*
Krieger, Henry, *Harvey Mudd College*
Krusemeyer, Mark, *Carleton College*
Kung, Sidney, *Jacksonville University*
Lagarias, Jeffrey, *AT&T Bell Labs, Murray Hill, NJ*
Lange, Lester H., *Moss Landing Laboratories, Moss Landing, CA*
Langer, John, *The Aerospace Corp., Los Angeles*
Larson, Dean, *Gonzaga University*
Larson, Suzanne, *Loyola Marymount University*
Lax, Peter D., *New York University-Courant*
Lay, David C., *University of Maryland, College Park*
Levine, Alan, *Franklin & Marshall College*
Lind, R. Bruce, *University of Puget Sound*
LoBello, Anthony J., *Allegheny College*
Long, Calvin T., *Washington State University*
Marcus, Marvin, *University of California, Santa Barbara*
Markham, Elizabeth, *St. Joseph's College*
Markley, Nelson, *University of Maryland, College Park*
Massell, Paul B., *U.S. Naval Academy, Annapolis, MD*
Maurer, Stephen B., *Swarthmore College*
May, Coy, *Towson State University*
McCleary, John, *Vassar College*
McDermot, Richard, *Allegheny College*
McDonald, Joanne M., *University of Regina*
McEachin, Raymond, *Indiana University-Purdue University, Fort Wayne*
Michaels, John G., *SUNY College at Brockport*
Monson, Barry, *University of New Brunswick*
Moore, Thomas L., *Grinnell College*
Morrison, John, *Towson State University*
Moser, William O.J., *McGill University*
Mugler, Dale, *University of Akron*
Nelsen, Roger B., *Lewis & Clark College*
Nievergelt, Yves, *Eastern Washington University*
Niven, Ivan, *University of Oregon*
Notz, William I., *Ohio State University*
Nutt, Michael D., *Acadia University*
Ostrom, T.G., *Pullman, WA*

Page, Warren, *New York City Technical College*
Pedersen, Jean, *Santa Clara University*
Peeples, Joanne V., *El Paso Community College*
Penn, Howard Lewis, *U.S. Naval Academy, Annapolis, MD*
Peterson, Bruce, *Middlebury College*
Pfiefer, Richard, *San Jose State University*
Plank, Donald L., *Stockton State College*
Quine, W.V., *Harvard University*
Ralston, Anthony, *SUNY at Buffalo*
Rebman, Kenneth R., *California State University, Hayward*
Reiter, Clifford, *Lafayette College*
Reznick, Bruce, *University of Illinois*
Richman, Dan, *Towson State University*
Rigby, John F., *University of Wales, College of Cardiff*
Roberts, Wayne, *Macalester College*
Robertson, Jack M., *Washington State University*
Rose, William R., *Towson State University*
Rosen, Kenneth, *AT&T Bell Labs, Middletown, NJ*
Rosenstein, George M., *Franklin & Marshall College*
Ross, Peter, *Santa Clara University*
Rothman, Neal J., *Indiana University-Purdue University, Indianapolis*
Rubinstein, Moshe, *University of California, Los Angeles*
Saff, Edward, *University of South Florida*
Sagan, Hans, *North Carolina State University*
Sandefur, James T., *Georgetown University*
Sangwine-Yager, Jane, *St. Mary's College*
Schattschneider, Doris M., *Moravian College*
Schaumberger, Norman, *Hofstra University*
Schecter, Murray, *Lehigh University*
Schiller, Barry, *Rhode Island College*
Schultz, Henry, *3D Systems, Saugus, CA*
Schuster, Seymour, *Carleton College*
Schwenk, Allen, *Western Michigan University*
Segal, Sanford, *University of Rochester*
Simon, Julie, *Middlebury College*
Singmaster, David, *South Bank Polytechnic*
Smith, Bryan, *University of Puget Sound*
Smith, David A., *Duke University*
Snell, J. Laurie, *Dartmouth College*
Sohrab, Houshang, *Towson State University*
Spanier, Jerome, *Claremont Graduate School*
Stark, Richard, *St. Mary's College*
Stern, Leonid, *Towson State University*
Stone, Alexander P., *University of New Mexico*
Stone, Arthur H., *Northeastern University*
Strang, Gilbert, *Massachusetts Institute of Technology*
Stromquist, Walter, *Daniel H. Wagner Assoc., Berwyn, PA*
Stroyan, Keith, *University of Iowa*
Taylor, Alan, *Union College*
Teles, Elizabeth J., *Montgomery College*
Terrell, Maria, *Cornell University*
Thuente, David J., *Indiana University-Purdue University, Fort Wayne*
Tucker, Alan, *SUNY at Stony Brook*
Tucker, Thomas W., *Colgate University*
Van de Wetering, R.L., *San Diego State University*
Vanden Eynden, Charles L., *Illinois State University*
Wagner, Anne, *Towson State University*
Wakin, Shirley, *Southern Connecticut State University*
Ward, Robert L., *Laurel, MD*
Warner, C. Robert, *University of Maryland, College Park*
Wilker, John B., *University of Toronto*
Willard, Stephen, *University of Alberta*
Williams, Robert C., *Alfred University*
Winter, June, *Marymount University*
Witte, David Riley, *Williams College*
Wood, Thomas, *Lafayette College*
Woodard, Mark, *Furman University*
Zerger, Monte J., *Adams State College*
Zimmerman, Jay, *Towson State University*
Zorn, Paul, *St. Olaf College*

The Seventh International Congress on Mathematical Education (ICME-7) will be held at Universite Laval in the city of Québec, Canada, Aug. 17 to 23, 1992. Among the programs featured is a miniconference "Computers and Calculators in the Curriculum."
For more information, contact: Congres ICME-7 Congress,
Universite Laval, Québec, QC, Canada G1K 7P4.
e-mail: ICME-7@VM1.ULAVAL.CA or fax: (418)656-2000.

Statement of Ownership, Management and Circulation
(Required by 39 U.S.C. 3685)

1A. Title of Publication	1B. Publication No.	2. Date of Filing
MATHEMATICS MAGAZINE	0 0 2 5 5 7 0 X	Oct. 11, 1991

3. Frequency of Issue	3A. No. of Issues Published Annually	3B. Annual Subscription Price
bi-monthly except July and August	Five	Library $68 Indiv. Member $16

4. Complete Mailing Address of Known Office of Publication *(Street, City, County, State and ZIP+4 Code) (Not printers)*

1529 Eighteenth St., N.W., Washington, D.C. 20036

5. Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*

1529 Eighteenth St., N.W., Washington, D.C. 20036

6. Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor *(This item MUST NOT be blank)*

Publisher *(Name and Complete Mailing Address)*

The Mathematical Association of America, 1529 Eighteenth St., N.W., Washington, D.C. 20036

Editor *(Name and Complete Mailing Address)*

Martha J. Siegel, Towson State Univ., Dept. of Mathematics, Baltimore, MD 21204

Managing Editor *(Name and Complete Mailing Address)*

Harry Waldman, MAA, 1529 Eighteenth St., N.W., Washington, D.C. 20036

7. Owner *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

Full Name	Complete Mailing Address
The Mathematical Association of America, Inc.	1529 Eighteenth St., N.W., Washington, D.C. 20036

8. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities *(If there are none, so state)*

Full Name	Complete Mailing Address
None	

9. For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 424.12 only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] Has Not Changed During Preceding 12 Months
(2) [ ] Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement.)

10. Extent and Nature of Circulation *(See instructions on reverse side)*	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A. Total No. Copies *(Net Press Run)*	20,536	19,926
B. Paid and/or Requested Circulation 1. Sales through dealers and carriers, street vendors and counter sales	0	0
2. Mail Subscription *(Paid and/or requested)*	16,780	15,322
C. Total Paid and/or Requested Circulation *(Sum of 10B1 and 10B2)*	16,780	15,322
D. Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	712	654
E. Total Distribution *(Sum of C and D)*	17,492	15,976
F. Copies Not Distributed 1. Office use, left over, unaccounted, spoiled after printing	3,044	3,950
2. Return from News Agents	0	0
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	20,536	19,926

11. **I certify that the statements made by me above are correct and complete**

Signature and Title of Editor, Publisher, Business Manager, or Owner
Kay Lamont, Membership Manager

PS Form 3526, January 1991 *(See instructions on reverse)*

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

INDEX TO VOLUME 64

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, D.C. 20036