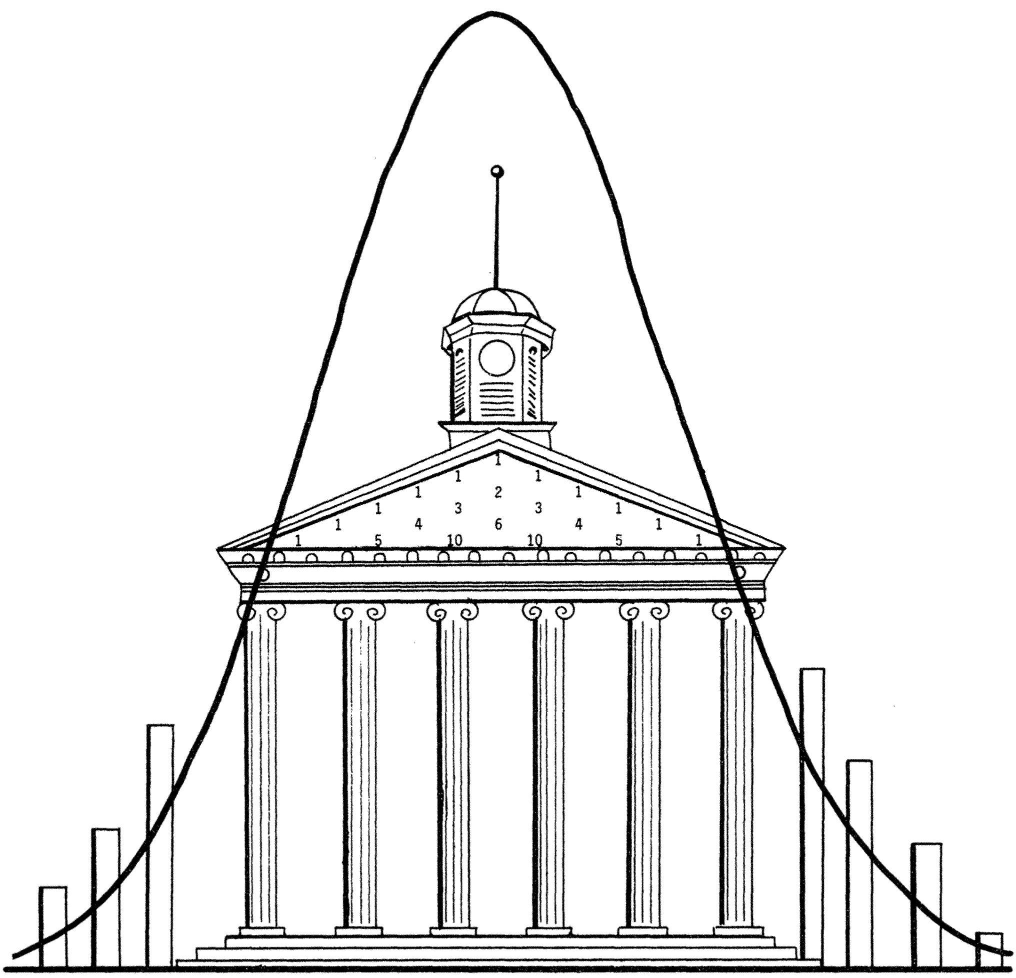
MATHEMATICS MAGAZINE



Vol. 56 No. 2
March, 1983

STATISTICS AND THE LAW • APOLLONIUS' PROBLEM
INCOMMENSURABILITY PROOFS • WROŃSKI'S CANONS

TWO-YEAR COLLEGE MATHEMATICS READINGS

Edited by Warren Page

viii + 304 pages. Paperbound
List $19.50   MAA Member: $14.50

This work surveys the two-year college mathematics landscape. Stops are made at points of interest that cover areas from algebra to recreational mathematics.

There are many things of interest for everyone in this book. Some are expository articles that introduce the reader to new advances in mathematics. Other articles can be used to enrich course work and to stimulate both the student's and instructor's mathematical intellect.

This book will serve as an excellent source of enrichment for the talented student. Teachers of undergraduates will find it is an excellent reference source that they will want to consult when preparing lectures.

Topics covered are:

- **Algebra**
- **Geometry**
- **Number Theory**
- **Calculus**
- **Probability and Statistics**
- **Calculators and Computers**
- **Mathematics Education**
- **Recreational Mathematics**

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

THE MATHEMATICAL ASSOCIATION OF AMERICA

MATHEMATICS MAGAZINE

Volume 56, Number 2, March 1983
(ISSN—0025-570X)

COVER: Statistics and the law. Design by the editor; normal curve by Gary Eichelsdorfer.

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

EDITORIAL POLICY

Mathematics Magazine is a journal which aims to provide inviting, informal mathematical exposition. Manuscripts accepted for publication in the *Magazine* should be written in a clear and lively expository style and stocked with appropriate examples and graphics. Our advice to authors is: say something new in an appealing way or say something old in a refreshing way. The *Magazine* is not a research journal and so the style, quality, and level of articles submitted for publication should realistically permit their use to supplement undergraduate courses. The editor invites manuscripts that provide insight into the history and application of mathematics, that point out interrelationships between several branches of mathematics and that illustrate the fun of doing mathematics.

Authors planning to submit manuscripts should read the full statement of editorial policy which appears in this *Magazine*, Vol. 54, pp. 44–45, and is also 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.

New manuscripts should be sent to: Doris Schattschneider, Editor, Mathematics Magazine, Moravian College, Bethlehem, PA 18018. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine*. They should be typewritten and double spaced on $8\frac{1}{2}$ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added.

AUTHORS

Mary W. Gray ("Statistics and the Law") is professor of mathematics, statistics and computer science at American University, Washington, DC. She has a Ph.D. in mathematics from the University of Kansas, a law degree from American University and is a member of the District of Columbia Bar. She has taught courses in statistics for lawyers and frequently serves as an expert witness in litigation involving the use of statistics. Her current research interests are in applied statistics and computer reliability and security.

E. J. Barbeau, ("Incommensurability Proofs: A Pattern that Peters Out") a native of Toronto and a Ph.D. graduate of the University of Newcastle-upon-Tyne (England), taught at the University of Western Ontario before moving to the University of Toronto in 1967, where he is currently associate professor. His research has been in functional analysis, history of analysis and number theory, and he has been involved with mathematical competitions at the High School and University levels. Currently, he is chairman of the Education Committee of the Canadian Mathematical Society. He is also an organist.

ILLUSTRATIONS

Vic Norton cartooned the trials of statistics and justice on pp. 69, 71, 75.

William McWorter, Jr. produced the computer-drawn figure p. 119.

All other illustrations were provided by the authors.

BUSINESS INFORMATION. Mathematics Magazine is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. The annual subscription price for Mathematics Magazine to an individual member of the Association is \$10, included as part of the annual dues of \$40. Students receive a 50% discount. Bulk subscriptions (5 or more copies to a single address) are available to colleges and universities for distribution to undergraduate students at a 35% discount. The library subscription price is \$25.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N. W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.

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

ARTICLES

Statistics and the Law

Intuitive views of evidence may be altered by mathematical analysis.

MARY W. GRAY
American University
Washington, DC 20016

Mathematical models have found their way into our legal system as into most aspects of present-day life. For the most part the models to which we have recourse in the law are probabilistic, not deterministic. Great trial lawyers are masters of the subtleties of probability. Leibniz once remarked that jurists have practiced best of all mankind the art of logic in regard to contingencies [**21**]. Indeed, central to the question of the admissibility of evidence is its *relevance*, which the Federal Rules of Evidence define in terms of probability: Evidence is relevant if it has "[a]ny tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence" [**14**]. However, as we shall show, the intuition of attorneys and judges may need the guidance of an analysis relying on various statistical techniques. Some have recognized this need. Mr. Justice Holmes, writing in 1897, said: "For the rational study of the law,... the man of the future is the man of statistics" [**20**].

Courts have greeted offers of guidance in statistics with varying degrees of receptivity and skepticism. For example, in *Alabama v. United States* [**1**], the court said, "Statistics often tell much and Courts listen." On the other hand, in *People v. Collins* [**25**], the court cautioned, "Mathematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not cast a spell over him."

There are three general areas in which statistics is of assistance to the law: (1) in evidence, (2) as a model of the trial process itself, and (3) for support purposes. Poisson pioneered in the modeling of the trial process with his analysis of jury verdicts [**27**]; today some visionaries, critical of the delays and vagaries of our judicial system, propose that we replace judges or juries by computers properly programmed with all precedents, together with elaborately constructed probability-based decision rules. In support for the judicial process, statistical analyses have been done to decide whether pretrial conferences will reduce trial duration [**15**], whether pretrial detention increases the probability of conviction [**23**], or whether capital punishment is a deterrent to murder [**6**], [**13**].

It is with the use of *statistics as evidence* that we are concerned here. Some applications seem trivial; others have caused people to be imprisoned.

One example of the former is interesting because it sheds some light on what probability one judge attached to "beyond a reasonable doubt." "Beyond a reasonable doubt" is the standard of proof for conviction in a criminal matter; most civil matters are decided on the basis of the preponderance of the evidence (although some cases such as fraud require the stricter standard of "clear and convincing evidence"). In a Swedish case [**24**] the defendant was charged with overtime parking at a place that had a time limit but no parking meter. To establish that his car had not moved during that time interval, the policeman had recorded the position of the tire air-valves on one side of the defendant's car. The record was kept in the manner that pilots indicate directions: the front-wheel valve was recorded pointing at one o'clock, the rear-wheel valve at eight o'clock. At the officer's second visit to the parked car—after the allowed time had expired—the same positions were recorded.

The driver claimed that he had been away and when he returned to the same parking place by chance the valves had come to rest in approximately the same positions. The court accepted the defense—calculating the probability of this outcome as $1/12 \times 1/12 = 1/144$—large enough, it said, to establish reasonable doubt. It went on to say that had all four valve positions been recorded and found to be the same, the probability of $1/12^4 = 1/20{,}736$ would have been small enough to be accepted as proof beyond a reasonable doubt. The problem with the method of calculating the probabilities used here is that while car wheels do not rotate completely parallel—for example when the car turns—neither are their rotations completely independent.

Identification evidence

The lawyer who wants to use statistics or the statistician who serves as an expert witness in a trial faces many problems in making the statistical evidence understandable and credible. Even the best of mathematicians may not be successful in a presentation to a tribunal. In the celebrated Dreyfus case [**19**] in which Alfred Dreyfus was tried for treason, among other charges it was alleged that letters Dreyfus wrote to various people contained cipher messages. This assertion was based on an analysis of the frequency of occurrence of letters of the alphabet in these messages which showed that symbols of the alphabet did not occur in Dreyfus' letters in the proportion in which they are known to occur in average French prose. Henri Poincaré tried to convince the hostile tribunal that among all possible proportions in which the letters might appear, any particular exact proportion—even the most probable—was very unlikely to occur in any given piece of writing. He failed. Anyone who has tried to convince a skeptic that even though the most probable number of heads in 100 tosses of a fair coin is 50, yet the probability of getting exactly 50 heads is less than 1/10 will appreciate Poincaré's difficulties. After all,

$$p(50) = \binom{100}{50}(.5)^{50}(.5)^{50} < .10.$$

Painlevé (less distinguished as a mathematician, but destined to become premier of France) was no more successful than Poincaré, in spite of an impassioned declaration: "Give me the works of Racine and I will show you that he, too, by your foolish tests is a traitor, for the works of Racine, like the letters of Dreyfus, do not show the most probable distribution." After two trials in which Dreyfus was found guilty, as part of the consideration of an appeal the Court of Cassation appointed a commission of three mathematicians, Poincaré, Darboux, and Appell, to report on the use of probability at the trials, a report which may have made some small contribution to the exoneration of Dreyfus [**28**].

Another legal case which involved a well-known mathematician concerned the authenticity of the signature of Sylvia Ann Howland on what her niece claimed was her second (and true) will [**29**]. Professor Benjamin Peirce of Harvard was called to give expert testimony. His calculations were based upon the observed and accurately tabulated coincidences in 50 genuine signatures and upon the assumption that each downward stroke is independent of every other. He found that characteristic lines coincided about once in five times, so the probability of coincidence was estimated to be 1/5.

In comparing the disputed signature with one known to be genuine on a first will, Peirce found thirty complete coincidences and so computed the probability of the two signatures being so much alike (were the second not traced from the first) to be one in 5^{30} (approximately 10^{21}). No matter how impressive Peirce's calculations might have been, the decision of the appeal court to throw

out the will turned on whether another witness was competent to testify, not on Peirce's testimony.

In another case, that of millionaire W. M. Rice, the signature on his will was disputed, and the will was declared a forgery on the basis of probability evidence. As a result, the fortune of Rice went to found Rice Institute.

Apparently judges put less faith in probability when it is applied to mechanical objects than to handwriting. A leading case, which for many years provided persuasive precedent for the suppression of probability evidence, involved a typewriter with peculiar characteristics. The defendant in *People v. Risley* [**26**], who was an attorney, was accused of abstracting an affidavit from the file of the court clerk, altering it, and replacing it. The six letters used in the alteration showed eleven peculiarities of type matched by the corresponding letters on the defendant's typewriter. The mathematics professor who testified came up with a figure of 1/4,000,000,000, although exactly what that was meant to represent or how it was computed is unclear in the appellate court's opinion. However, the conviction obtained in the trial court was overturned by the appeal court on the grounds (1) that the mathematician could not be qualified as an expert in typewriting and (2) that he did a snow job on the jury: "The minds of the jury were thus diverted from the actual facts in issue and led into a maze of abstract theory and speculation." Sherlock Holmes fans may recall that he once solved a typewriter case [**12**] by a greater faith in the abstractions of probability than had the appeal court in *Risley*.

People v. Collins

The most famous case cited for the misuse of probability, and that which has stirred the most mathematical controversy, is *People v. Collins* [**25**]. After an elderly person was mugged in an alley of a Los Angeles suburb, a couple was seen running from the alley. They were described as a black man with a beard and a mustache and a blond girl with her hair in a ponytail. According to witnesses they drove off in a partly yellow car.

Malcolm and Janet Collins were later arrested. He was black, with evidence of recently having had a beard and mustache (although he was clean-shaven when arrested). She was blond and customarily wore her hair in a ponytail. They had a partly yellow Lincoln.

The prosecution in their trial called a professor of mathematics as a witness. "Just suppose," the prosecutor said, "that we could assign the following conservative probabilities to the characteristics noted by the witnesses:

partly yellow automobile	1/10
man with mustache	1/4
girl with ponytail	1/10
girl with blond hair	1/3
black man with beard	1/10
interracial couple in car	1/1000.

"Now," he continued, "is it not true that if we have independent events we find the probability that they all occur by multiplying the individual probabilities?" The poor professor tried to protest that these probabilities were far from independent, but was directed: "Just answer the question." "Yes," he had to admit, "one does multiply." "And the product of these probabilities is 1/12,000,000, is it not?" No arguing with that, and the witness was dismissed, with the dazed defense befuddled by all this high-powered mathematics.

Later in the trial the prosecutor claimed that inasmuch as there were 12,000,000 people in the Los Angeles metropolitan area, and his probability estimates were conservative, the chances of there being another couple like the Collinses in the area must be one in a billion and thus he had "mathematical proof" of their guilt. Unfortunately the jury convicted on little evidence other than this misuse of statistics.

On Malcolm Collins' appeal, the California Supreme Court reversed the conviction, but only after he had spent considerable time in prison. The court cited two grounds for finding that the mathematical testimony was improper:

(1) no foundation for the hypotheses had been laid, i.e., the probability estimates were pure speculation;

(2) the use made of the probability testimony was erroneous and distracted the jury from its proper factfinding function. The court said, "Undoubtedly the jurors were unduly impressed by the mystique of the mathematical demonstration but were unable to assess its relevancy or value."

The court's own probability analysis was quite simple [**25**]. Suppose we consider the Los Angeles population to be a sample from a population where the frequency of the characteristics is as given; suppose, moreover, that there is one couple with the given characteristics, namely the couple who committed the crime. What is the probability that there is another couple with the same listed characteristics? Note that this is not the same probability as the probability that the Collinses, having the same characteristics, are the couple who perpetrated the crime.

To continue with the *Collins* court's analysis: if the probability of a randomly chosen couple having the six listed characteristics is 1/12,000,000 and if we know that there is one such couple, then the probability of there being more than one is 42%. Because the probability is small and the size of the sample large, we use a Poisson distribution to compute this figure as follows:

$$p(\text{more than one given at least one}) = \frac{p(\text{more than one})}{p(\text{at least one})}$$

$$p(\geqslant 1) = 1 - p(0) + 1 - 1/e = .63$$

$$p(>1) = 1 - p(0) - p(1) = 1 - 1/e - 1/e = .26$$

$$\frac{p(>1)}{p(\geqslant 1)} = .42.$$

This calculation, along with the other difficulties with the role statistics had played in the conviction, convinced the appellate court to reverse the trial court. That there are other difficulties in the original analysis is apparent; for example, we are speaking of couples, not individuals; there may have been non-Los Angeles individuals in the area; whether it is appropriate to look at the 12,000,000 population of the area as a 12,000,000 person sample is questionable; we still do not have a basis for the original estimates nor are their probabilities independent.

A Bayesian approach

Commentators have proposed other approaches to the *Collins* case. In particular, one could reword the traits of the muggers as described by the witnesses so that the probabilities would be independent and calculate the probability of the Collinses being the muggers, given that both have the same traits. Another possibility is to use a Bayesian approach, which would require making an estimate of the probability of the Collinses' being the muggers, excluding the similarity evidence, and then calculating a posterior probability using the information as to the frequency of the given characteristics in the population.

Advocates of the Bayesian approach would proceed as follows. Suppose we are confronted with a knife with a palm print on it which is similar to the defendant's palm print. We want to know the probability that the print is that of the defendant. Suppose also that we have some information apart from the characteristics of the print itself which enables us to estimate the probability that it is that of the defendant. For example, suppose she was seen near the scene of the crime, she was known to hate the victim or she has an apparently ironclad alibi. This information gives us an intuitive prior probability that the print is hers.

Let G be the event that the defendant's print is on a knife used in a murder (not the event of her guilt) and let H be the event that a palm print similar to the defendant's is found on the knife. For simplicity assume that $p(H|G)=1$ and that $p(H|\text{ not } G)$ is the frequency of the characteristics of the palm print among palm prints in the general population. Then

$$p(G|H)=\frac{p(G)p(H|G)}{p(G)p(H|G)+p(\text{not } G)p(H|\text{ not } G)}.$$

Considerable modification of the prior probabilities can occur by using the probability evidence about the print's characteristics (see TABLE 1).

Frequency of characteristics $P(H\|\text{ not } G)$	Prior probabilities $p(G)$ .01	.25	.50
.50	.019	.400	.666
.10	.091	.769	.909
.01	.502	.970	.990
.001	.909	.997	.999

TABLE 1. **Posterior Probabilities $p(G|H)$.**

For example, if we previously thought that there was only one chance in one hundred that the print belonged to the defendant, knowledge that the print characteristics occur only once in a thousand times causes us to raise our estimate to 91 chances in 100.

The argument advanced for this analysis is primarily that it focuses on the real question: how much more likely is the defendant to have left the print than someone from the general population? As you may imagine, not everyone believes that this sort of analysis is an improvement, or even proper. The likelihood that a jury will be able to understand and interpret the model does not seem to some to be greater than the likelihood that they will be able properly to deal with an unadorned estimate of the frequency of the characteristics. (This also ignores the question of the propriety of asking the jury to make the preliminary judgments which would be required in order to get the prior probability before all the evidence was in.) In any event, there is no present likelihood of Bayesian analysis taking over the criminal courts [**31**].

Discrimination cases

In discrimination cases, statistical evidence can show that a hypothesis of random selection, that is, of no discrimination, is so improbable as to make it likely that some other process must have been at work. But it cannot *prove* that there was another factor, and it certainly cannot show that the factor was discrimination.

The use of probability in discrimination cases has, in many instances, been a straightforward application of a binomial model. However, the courts have taken many years to acknowledge the role of probability in helping them find evidence of discrimination. The original cases (1880–1960) dealt with discrimination in the selection of a jury. In a typical situation, a black was indicted by a grand jury and convicted by a petit jury from which blacks were excluded to varying degrees. The defendant sought to get the federal courts to overturn a state court conviction on the grounds of abridgement of his right to have a jury of his peers, a right extended to state court cases by the Fourteenth Amendment. Later variations involved discrimination against Chicanos or women (although the Court for many years felt it quite proper to limit the jury service of women by various devices), or the assertion by a white of the violation of his rights by the exclusion of blacks from the jury which tried him.

The cases in the 1880's presented a fairly straightforward matter; there were state laws which excluded all blacks or the state admitted the fact of deliberate exclusion. Soon the practice of discrimination by local governments became more subtle and it was only in practice, not in law or in announced policy, that blacks were excluded. However, as long as there was total exclusion of blacks, the Supreme Court generally found discrimination and overturned convictions. Total exclusion of blacks from juries was replaced by tokenism and local authorities were careful to include some blacks on juries. In response to this, the Supreme Court took the view that one should look at the difference between the percentage of blacks on the grand jury or petit jury panel and the percentage in the local population or on tax or voter lists. For example, in *Swain v. Alabama* [**30**] blacks constituted 26% of the adult population of Talladega County and about 16% of the grand and petit jury venires, the groups from which juries are actually selected. Ten percent difference, said the Court, cannot be taken to show discrimination. From the venire, a petit jury is chosen for a particular trial. Both prosecution and defense can challenge jurors for cause or, in a limited number of cases, peremptorily (without stated cause). In Talladega County blacks had never actually *served* on petit juries, but Swain could not show that in each case the prosecution had been responsible for striking them (the prosecution had struck the six eligible blacks on Swain's panel).

On the other hand, in *Avery v. Georgia* [**2**] the defendant was convicted by a jury selected from a panel of 60 where the names of the 60 were chosen from a box containing tickets with the names of persons on the jury roll. Although the selection from the box was supposedly at random, none of the 5% who were blacks were selected. (It also happened that the names of blacks were printed on yellow paper and the names of whites on white paper.) Thus, although the reduction was only from 5% to 0%, the end result was *no* blacks on the panel. In the successful challenge to the conviction, the yellow slips were thought by the Court to be a "plus" factor pointing to discrimination. Mr. Justice Frankfurter said, "The mind of justice, not merely its eyes, would have to be blind to attribute such an occurrence to mere fortuity."

	% Minorities in pool	% Minorities on jury panels	Probability
Swain v. Alabama	26%	16%	10^{-8}
Avery v. Georgia	5%	0%	4.6×10^{-2}
Whitus v. Georgia	27%	8%	6×10^{-6}
Castaneda v. Partida	79%	39%	10^{-140}

TABLE 2

Let us see how fortune would view the two cases discussed as well as two others (see TABLE 2). The probabilities in this table represent the probabilities of the results as extreme or more extreme than those which occurred, using a binomial distribution, since the pool in each case was essentially infinite.

In *Whitus v. Georgia* [**33**], jury panels were chosen from tax rolls in which 27% were black. However, all blacks on the rolls had a special designation (c) attached to their names. The panels selected were 8% black. The Court took judicial notice that the likelihood of such an outcome, were the selection at random, was 6×10^{-6}, but specifically said that its decision did not depend on the probability analysis, since the (c) designation was considered crucial. Only in *Castaneda v. Partida* [**8**] did the Court acknowledge its reliance on probability. In Texas, "key men" in the community were responsible for selecting the jurors. Even though most of the "key men" were Chicanos, in a county 79% Chicano the jury venires were only 39% Chicano, a result whose probability is 1 in 10^{140} if the selection were random. On this basis the Court overturned the trial court's conviction.

One of the most interesting of the losing challenges to the composition of juries was the case of Dr. Spock and the vanishing women jurors [**34**]. In 1968 Dr. Benjamin Spock was tried for conspiracy to violate the Military Services Act of 1967 [**32**]. Under the method of jury selection then in effect in the U. S. District Court for the District of Massachusetts, names were selected from lists of all residents age 21 or over (of whom 56% were women) and questionnaires were sent by the clerk of the court to those selected. After eliminating those disqualified by statute from jury service, the clerk placed batches of 300 names into a central jury box; from this box, the clerk, prior to the scheduled jury trial, drew a venire. Those whose names were drawn were ordered to appear in court on the day of the trial.

The situation was as follows:

	Percentage of women
Residents over 21	56%
Average for all venires	29%
Average of trial judge's venires	14.6%
Venire for Spock trial	9%
Spock jury	0%.

Of the 100 who appeared for jury duty for Dr. Spock's trial, only nine were women. The defense challenged the composition of this array and lost. The jury eventually selected from this venire had no women members. Dr. Spock was convicted, but his conviction was subsequently reversed, on First Amendment grounds, not on the grounds of improper jury selection procedures.

The venires for the trial judge in Dr. Spock's case averaged 14.6% women, an average half that of the venires of the judge's six colleagues on the same court. That the distribution in the case of Dr. Spock's judge should be so different from that of his colleagues if the venires were drawn in the same random fashion from the central jury box is very unlikely. If we compare the distribution of the venires of the *Spock* judge with what would be expected if he drew his venires in the same way as his colleagues, using a χ^2 test we obtain a probability of 1 in 10^{18} that the method of selection was the same.

One may wonder why anyone would want women excluded from a jury; or, for that matter, why Dr. Spock's defense wanted them included, other than because of an abstract belief in the integrity of the operation of the jury system. One might conjecture that it was because a generation of women relied on Dr. Spock's baby book and hence would be reluctant to convict such an authority figure. In fact, the defense was relying on the statistics of a Gallup poll which showed that 50% of the men, but only 32% of the women, characterized themselves as "hawks," wanting to increase the military effort in Vietnam. Hawks were, not surprisingly, thought to be more likely to be hostile to Dr. Spock.

Model building

Let us look briefly at the legal issue which underlies the question of the composition of either a jury or an employer's workforce. The Constitution guarantees a "jury of one's peers," which has been interpreted to mean that the panel must be selected in such a way that each eligible person has an equal chance of being selected. (Note that we speak of jury *panels* to eliminate the thorny issue of peremptory challenges in the selection of the actual trial jury from the panel.) Similarly, antidiscrimination laws guarantee that similarly qualified persons of different races or genders should have an equal chance of being employed or promoted. A defendant is not entitled to a jury which reflects the population nor is an employer required to have a workforce which reflects the composition of the available labor pool. When the selection of jurymen and jurywomen or of employees includes very few blacks or very few women, this may be taken as *prima facie* indication of discrimination when such an outcome is very unlikely if the selection were random from among those eligible. If a plaintiff establishes a *prima facie* case, the burden shifts to the defendant to rebut the evidence. For example, the defendant might show that in fact those disproportionately excluded were less qualified.

Note that our concern is with the probability that a result occurs by chance. Thus a jury panel which always shows the same percentage of blacks (a percentage closely matching the percentage of blacks in the pool) is just as suspect as one which excludes blacks. For to get such an outcome time after time, something other than random selection must have played a role.

In Dallas County, Texas, several convictions were overturned because of jury discrimination. After the reversals, 17 blacks were included among the 252 members of grand juries. One of these juries indicted a black man named Cassell [7]. The discrepancy between the $17/252 = 6.7\%$ and the 15.5% black population was accepted by the Court because blacks constituted only 6.5% of the poll-tax payers, and payment of the poll-tax was prerequisite for jury service. A probability analysis of the aggregate figures would give no cause for suspicion of discrimination since the percentage of blacks on the juries (6.7%) is larger than the percentage of blacks in the eligible population (6.5%). However, of the 21 juries made up from 252 jurors, 17 had just one black, the others none. The Court held that a limitation to one black per jury would be unconstitutional since "jurymen should be selected as individuals... and not as members of a race." However, no analysis was done of how likely it was that they were selected as individuals, that is, of how likely it was that the given figures would have resulted by chance if no other factor played a role. Had the Court analyzed the statistics on distribution, they would have found

	Expected number	Observed number
Juries with no blacks	9.14	4
Juries with one black	7.87	17
Juries with more than one black	3.99	0.

The expected numbers are computed using the proportion of blacks on the combined juries, 6.7%.

We find

$$\chi^2 = \sum \frac{(\text{expected number} - \text{observed number})^2}{\text{expected number}} = 17.47.$$

The probability of such a large χ^2 value is less than .001, making it likely that a limitation on blacks or some other nonrandom factor was in operation. In fact, the *Cassell* Court vacated the conviction because of the testimony of the jury commissioners that (1) they took only people whom they knew for jurors and (2) they did not know any blacks.

The chi-square analysis in *Cassell* is an example of a nonintuitive way of looking at the problem. What might seem at first a very equitable distribution of blacks on the juries, a distribution which random selection might well produce, becomes instead an indication of a deliberate scheme of limitation when a different model is utilized. The issue of limitation to one black at most per jury is more important than it might seem, for studies have shown that for a minority view to have any effect it is almost essential that the minority not be a minority of one [**35**].

This example points up the fact that the usual source of error in the use of statistics is not some gross miscalculation, but rather the specification of the wrong model because some relevant fact has been left out of the model employed. Sometimes finding an appropriate model is not all that easy.

Another example indicates the difficulty of formulating models. A stockbroker's computerized access to instant stock market quotations on his dial-up computer terminal was interrupted for three days by a phone company error in disconnecting the line. The other phones continued to operate so that some business was transacted; on the other hand, many customers would not buy or sell if they could not get an instant quote. The question was how to assess damages for this lost business. Ma Bell wanted merely to refund $1/10$ of the monthly phone service charge, but the broker wanted compensation for lost business. Thus the question became: how much business was lost? It seemed reasonable to look at records for comparable days. But then the question became complicated: how are they to be comparable? Same month, same days of the week, same weather, same gross market transactions, same trend in the market? In fact, it was difficult to find a pattern on which to base a model; for example, gross transactions and the transactions of the particular broker appeared to bear no relation to one another. No matter what sort of average was computed the standard deviation was very large, which put the broker's loss on the phoneless days well within chance variation. The best model, from the point of view of minimizing the standard deviation, came from taking a random sample of three-day periods and computing the average sales for the three days. The sales for the days of interrupted service were then compared with this average; the difference was somewhat more than one standard deviation. The case was settled out of court on the basis of estimates from this method, but there was no opportunity to convince a judge of the merits of the model.

Discrimination in employment and education

Many employment and education cases are very similar to the jury discrimination cases, simply an application of a binomial model. For example, close on the heels of *Castaneda* (p. 73), but after many lower court decisions on employment discrimination which relied on statistics, followed *Hazelwood School District v. United States* [**18**]. The Court computed the expected number of black teachers, compared that with the observed number, noted a variation by five or six standard deviations and quoted from *Castaneda*: "As a general rule for such large samples if the difference between the expected value and the observed is greater than two or three standard deviations, the hypothesis of no discrimination would be suspect."

Black teachers

	Expected number	Observed number	Difference
1972–73	63	16	6 standard deviations
1973–74	70	22	5 standard deviations

A potential case which never got to the litigation stage again involves a chi-square test; here the question was one of which model is appropriate for the situation [**4**]. This concerns graduate admissions at the University of California, Berkeley, in the early seventies. We make the assumption that in any given discipline, male and female applicants do not differ in the distribution of legitimate criteria for admission. TABLE 3 shows the aggregate data.

Applicants	Outcome: Observed		Outcome: Expected		Difference	
	Admit	Deny	Admit	Deny	Admit	Deny
Men	3738	4704	3460.7	4981.3	277.3	−277.3
Women	1494	2827	1771.3	2549.7	−277.3	277.3

TABLE 3

Approximately 44% of the men and 35% of the women were admitted. We see a shortfall of 277 women. Who is responsible? A look at the individual departments showed that sixteen departments had no women applicants or denied admission to no one. Among the remaining 85 departments or interdepartmental graduate majors, four had a "bias" against the admission of women in which the probability of that occurring by chance is less than 5%, for a "deficit" of 26 women. Six departments were "biased" in the opposite direction for a "deficit" of 64 men. Net result: a deficit of 38 men instead of 277 women. What happened? Why don't the calculations add up? Would it have been better to construct 85 individual contingency tables and to compute chi-square for each and then to aggregate the statistics rather than aggregating the departments before computing the statistics?

There are several possible procedures for aggregation. One is due to Fisher [16]: Let

$$F = -2 \sum_{i=1}^{n} \ln p(T_i),$$

where $p(T_i)$ is the probability value of the test statistic (here χ^2) calculated for the ith experiment (department). The statistic F is referred to the upper tail of a chi-square distribution with $2n$ degrees of freedom ($n = 85$). The probability is 2.9% that there is bias in some direction. On the other hand, a slight variation in the computation produces a one-sided statistic to measure the hypothesis of "no bias" against "bias in favor of women." We find a probability of 85% that the result could occur under the hypothesis of "no bias." What went wrong? These calculations also don't add up.

We have been looking for a link between the sex of the applicant and the decision to admit. In fact, there is a prior linkage: that between the sex of the applicant and the department to which admission is sought. For example, 2/3 of the applicants to English are women, but only 2% of the applicants to mechanical engineering are women. In fact, if we construct a 2×85 contingency table linking departments and sex of applicants we get $\chi^2 = 3027$, and the probability of obtaining a chi-square that large or larger by chance is approximately zero. So the distribution of women applicants across departments is far from uniform. These statistics are important because of the fact that not all departments are equally easy to enter. For the 2×85 table linking departments and admission rates, $\chi^2 = 2121$, and the probability that the admission rate is the same in all departments is approximately zero.

The explanation of the admissions data is that the proportion of women applicants tends to be high in departments that are hard to get into and low in those that are easy to get into. Moreover, the phenomenon is more pronounced in departments with large numbers of applicants. One could theorize as to cause and effect here, but the figures do explain the earlier apparent anomaly.

Let us go a bit further. If we accept that aggregating the data gives us misleading results, is disaggregation better? We are then looking at the results of 85 separate experiments. For example, the department most biased against women has a bias of sufficient magnitude that it would occur by chance in 69 of 100,000 cases, again using a χ^2 analysis. On the other hand, the probability of getting a department that is that biased (or more biased) against women by chance alone in 85 simultaneous trials is 57 in 1000. However, leaving aside the fact that even serious bias may go unrecognized in a totally disaggregated model, the other problem with disaggregation is that in many cases the departments are very small, making any sort of statistical analysis difficult.

Moreover, even if the department is really the decision-making unit, the campus is in some sense a unit. So it makes sense to ask: Is there a campus bias by sex in graduate admissions? What is difficult is to specify the model for measuring campus-wide bias. A suggested approach is to compute for each department separately the expected number of admissions by multiplying the number of women applicants by the department's admission rate and summing the expected numbers over the 85 departments.

1973	Expected female admittees	1432.9
	Observed female admittees	1493.0
	Difference	60.1

$$\chi^2 = 8.55 \qquad p = .003$$

Thus there appears to be a bias in favor of women for the 1973 year; the same approach in 1969 through 1972 shows no evidence of bias.

One might want to combine the above pooling approach with an examination of outliers (departments whose percentage of female admittees was far above or below the predicted value), again to apportion the blame, or to seek improvement, as the case may be. In this case, the data were aggregated over a five-year span to get large enough numbers to be meaningful. Again using a χ^2 analysis, it was found that two units showed significant bias in favor of women and two units showed significant bias in favor of men.

Regression and remedies

The use of regression analysis in various contexts has been more or less accepted by the courts. However, we address the question of what happens after the court has, in essence, endorsed the results of a regression analysis—in our case a regression analysis of salaries. There are two federal statutes addressing the issue of gender discrimination in compensation: the Equal Pay Act and Title VII of the Civil Rights Act of 1964. The first prescribes equal pay for equal work for men and women. Under its standards one is obliged to find a man and a woman who are or were doing essentially identical work under essentially identical conditions in the same establishment and who have the same seniority and show that they are compensated differently. This can obviously be difficult, especially in the case of professional employment, most particularly in the case of college and university faculty.

Title VII is more general: it prohibits discrimination on the basis of race, color, national origin and religion as well as gender; it covers compensation and terms and conditions of employment. Because of a provision called the Bennett Amendment, it was maintained by some that the only gender-based compensation claims which could be brought under Title VII were those covered by the Equal Pay Act, that is, those involving equal work. In the summer of 1981, in *County of Washington v. Gunther* [**10**], the Supreme Court held that Title VII's protection is broader, that claims of sex discrimination where the work was not equal could be heard. While the dimensions of the decision are not yet completely clear, it certainly is likely to make the use of regression analysis more widespread, although with what degree of success is difficult to predict.

What one does in regression analysis is to control those factors which would prevent a claim under the Equal Pay Act—differences in experience, skill, effort, and working conditions. Suppose we have convinced a court of the validity of a multiple regression study of salaries which shows evidence of possible sex discrimination; our analysis has withstood defense attacks on the choice of variables, composition of the group studied and any other challenges by the rival statistician. The plot of our data and the regression line are shown in FIGURE 1. We show only one independent variable for graphic purposes. Usually there will be several independent variables used in a salary regression; a study of faculty salaries might include such information as terminal degree, years of experience, and number of publications.

The regression line is "fitted" to the data points for men in the sense of least squares. That is, we find the straight line which minimizes the sum of the squares of the distances between the data points for men and points on the line with corresponding X-coordinates. The Y-coordinates of the data points give the actual salaries; the points on the line with the same X-coordinates are the predicted salaries and the differences are the residuals. We see that data points for both men and women are scattered around the regression line, which fits the data fairly well. The measure of this fit is called the correlation coefficient; its square tells us how much of the variation in salaries can be "explained" by the independent variables. It is important to understand that "explain" is used in a statistical sense; correlation is not causation. In FIGURE 1 we observe that there are more women below the line than above; if a regression line were computed using the data for them instead of that for the men, it would fall below the regression line for men.

We assume that the court has made a finding of discrimination in what is called the liability stage of the trial, and we want to propose a remedy. The immediate reaction might be to give more money to those women whose salaries lie below the regression line, or more conservatively, to examine their salaries to see whether they are "explained" by other factors not reflected in the independent variables. This approach has several difficulties. For example, what about the men whose salaries are below the line? They most certainly can make a claim of discriminatory treatment if no adjustment is made for them. (In fact, using this type of argument, male faculty at the University of Nebraska whose salaries fell below the level predicted by a formula resulting from a procedure similar to regression analysis succeeded in getting themselves raises [**5**].) The approach we have described which singles out women but not men is usually called "flagging." It will probably not withstand judicial scrutiny. Moreover, what about the women above the line? Perhaps were it not for discrimination their salaries would be even higher.

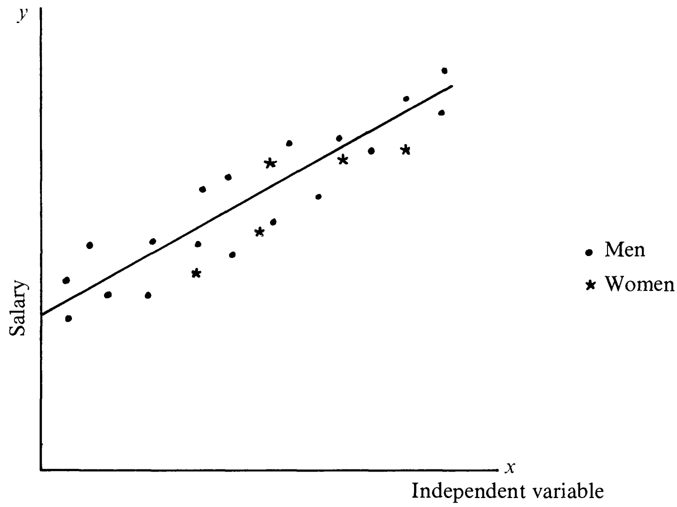


FIGURE 1

One possible alternative to flagging is to impose a statistical remedy since the disparity has been identified statistically [**17**]. There are various ways to do this, but the essence of the remedy would be to bring the regression line for the women up to that for the men. What we could do is to find the average residual for women, which under our assumption would be negative. We then add the absolute value of that amount to the salary of each woman. However, there is also a problem with this remedy, namely, we are establishing group rights rather than individual rights—the right of the group of women to be treated equally with the group of men. That is not the basis of statutory or constitutional guarantees of rights. The courts have recognized that once a finding of discrimination has been made, remedies may take race and sex into account in formulating a remedy in such a way as would give preference to members of one race or sex, impermissible otherwise. How the courts would view this particular proposed application of the principle remains untested.

Life expectancy

Our final example is the triumph of individual rights over group rights, although it is rarely viewed in this light. The issue is that of equality in pension benefits. Some pension plans, including the one in operation at a very large number of colleges and universities, differentiate on the basis of gender in paying out benefits. If a man and a woman have paid in the same amount, the woman, under one of the options available, will receive 15% less per month in pension benefits. Similarly, men either pay more or receive less in benefits in life insurance. Although many colleges and universities make up the difference on men's life insurance, virtually none do so for women's pensions.

In any event, the difference in pension benefits is "justified" on the basis of women's greater life expectancy. There are a lot of subsidiary issues. For example, is it sex or smoking which makes the difference? Are the life expectancies calculated for the appropriate group? Are the mortality tables used hopelessly out of date? Is the difference in life expectancies between men and women constant and is it properly computed? These are not our concerns here. We wish to cite a statistical argument which was very helpful in getting the Supreme Court, in *City of Los Angeles v. Manhart* [**9**], to declare that treating similarly situated men and women differently with respect to pension plans is a violation of Title VII.

Courts have long been receptive to actuarial evidence, notably in setting awards in wrongful death claims, but understandably, they have not been very sophisticated in understanding its basis. For example, it has been difficult to win acceptance of the concept that life expectancy can be computed in a variety of ways, but must be computed with respect to a particular group, although the group could be variously constituted. It seems difficult to get across the idea that one may have one life expectancy as an American, a longer one as a woman, a shorter one as a resident of Washington, D.C., a longer one as a nonsmoker, a shorter one as an overweight person—and so on—and that therefore to choose to compute for pension purposes life expectancy only on the basis of sex (and age) constitutes discrimination. It puts women at a disadvantage because of their sex, by treating every woman as the average woman.

The following concept [**3**] has proved useful in explaining to courts why paying benefits on the basis of sex is discriminatory, even when the above argument about life expectancies does not make much headway. Many traits are similarly distributed for men and women, but are shifted—for example, the ability to lift weights. Some women can lift more than some men, but on the average, women can lift less. For that reason, courts in employment discrimination cases have prohibited blanket restrictions in hiring. If a legitimate job requirement is the ability to lift 100 pounds, then a test must be given to applicants; one cannot just exclude all women because the average woman cannot lift 100 pounds.

Similarly, there is an overlap in the distribution of ages at death. But we cannot test for how long an individual will live; we could refine our prediction by looking at factors other than sex, but it would remain nothing more than a prediction. We do not know who will live a long life, and that is why pension plans insure against excessive longevity.

If we take a cohort of 1000 men age 65 and 1000 women age 65 and match their death ages, we would find that 84% can be matched up, leaving 8% of the population consisting of long-lived women unmatched by men and 8% of the population consisting of short-lived men unmatched by women. What the unlawful pension plans do is give every man a benefit for being of the same sex as the 8% who die early and every woman a burden for being of the same sex as the 8% who are longer-lived. The 84% of the population who are identically situated with respect to age at death are treated differently *solely* because of their sex; they receive less benefits (or pay more for the same benefits) without the dubious utility of a longer life. Thus, in the interest of equalizing benefits for men as a group and women as a group, individuals are treated differentially because of their sex. This, the Supreme Court said, violates the individual rights protections of Title VII [**9**].

A similar statistical argument can be made about automobile insurance rates, but at the moment there is no federal law prohibiting discrimination in insurance as Title VII prohibits discrimination in employment. Here, as in other areas of the law, the conclusions based on statistics may come into conflict with social policy. There is a strong policy against discrimination on the basis of race, so that statistics which show a higher accident rate for one race than for another are not used. In fact, one study in California showed that automobile accident rates were more strongly correlated with color of eyes and color of hair than with sex; no insurer has found it acceptable to differentiate in rates on the basis of color of eyes or hair. One would imagine that there would be a strong social policy in favor of basing insurance rates only on individual driving records, since unlike race, sex or age they are within the control of the individual, but that argument has yet to prevail over the actuarial evidence with respect to categorizations other than by race.

Conclusion

Powerful as statistics can be in guiding lawyers, judges, and juries along the course of justice, it is a weapon which must be employed with caution. In the law there is no opportunity to continue to search for counterexamples if a proposition cannot be proved; no possibility exists of a verdict that hedges—someone wins and someone loses. The stakes can be high, and the penalty severe when a model is misspecified or results subject to improper interpretation. A better understanding by statisticians of the underlying legal principles and by lawyers of the basic mathematical concepts could lead to the more productive employment of statistics in the law.

References

In the British and American common law, decisions are based on previous cases which serve as precedents. Since the system depends on the reporting of decisions and of the reasoning behind them, there are various compilations of these available. Federal court decisions and state court decisions for each state are reported separately, so that the system is quite complicated. Most law libraries have information available on how to find cases, given the citations in the form listed here.

[1] Alabama v. United States, 304 F. 2d 583 (5th Cir.), *aff'd per curiam*, 371 U. S. 37 (1962).
[2] Avery v. Georgia, 345 U. S. 559 (1952).
[3] B. Bergmann and M. W. Gray, Equality in retirement benefits, Civil Rights Digest, 8 (1975) 25–27.
[4] P. J. Bickel, E. A. Hammel and J. W. O'Connell, Sex bias in graduate admissions: data from Berkeley, Science, 187 (1975) 298–304.
[5] Board of Regents v. Dawes, 522 F. 2d 380 (8th Cir. 1975), *cert. denied*, 424 U. S. 914 (1976).
[6] W. J. Bowers and G. L. Pierce, The illusion of deterrence in Isaac Ehrlich's research on capital punishment, Yale Law J., 85 (1975) 187–207.
[7] Cassell v. Texas, 339 U. S. 282 (1950).
[8] Castaneda v. Partida, 430 U. S. 482 (1977).
[9] City of Los Angeles v. Manhart, 435 U. S. 702 (1978).
[10] County of Washington v. Gunther, 452 U. S. 161 (1981).
[11] A. D. Cullison, Identification by probabilities and trial by arithmetic (a lesson for beginners in how to be wrong with greater precision), Houston Law Rev., 6 (1969) 471–518.
[12] A. C. Doyle, A Case of Identity, Optimus Printing, London, 1895.
[13] I. Ehrlich, The deterrent effect of capital punishment, Amer. Economic Rev., 65 (1975) 397–417.
[14] Federal Rules of Evidence for United States Courts and Magistrates, West Publishing, St. Paul, 1975.
[15] M. O. Finkelstein, Quantitative Methods in Law, The Free Press, New York, 1978.
[16] R. A. Fisher, Statistical Methods for Research Workers, 13th ed., Hafner, New York, 1958.
[17] M. W. Gray and E. L. Scott, A statistical remedy for statistically identified discrimination, Academe, 66 (1980) 74–79.
[18] Hazelwood School District v. United States, 433 U. S. 299 (1977).
[19] E. W. Hendrick, The reality of mathematical processes, National Council of Teachers of Mathematics, 3rd yearbook, 1928.
[20] O. W. Holmes, The path of the law, Harvard Law Rev., 10 (1897) 457–469.
[21] G. W. Leibniz, Logical Papers, Oxford Univ. Press, Oxford, 1966.
[22] E. B. Mode, Probability and criminalistics, J. Amer. Statist. Assoc., 55 (1963) 628–640.
[23] Note, Preventive detention: an empirical analysis, Harv. Civil Rights-Civil Liberties Law Rev., 6 (1971) 289–396.
[24] Parkeringsfrågor, II. Tillorförlithitligheten av det s.k. klocksystemet för parkeringskontrol, Svensk juristidining, 47 (1962) 17–32.
[25] People v. Collins, 68 Ca. 2d 319, 438 P. 2d 33, 66 Ca. Rptr. 497 (1968).
[26] People v. Risley, 214 N. Y. 75, 108 N. E. 200 (1915).
[27] S. D. Poisson, Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités, Bachelier, Paris, 1837.
[28] Rapport de MM. les experts Darboux, Appell, et Poincaré, Les Documents Judiciaires de L'Affaire Dreyfus, La Revision du Procès de Rennes, Ligue Française pour la Défense des Droits de l'Homme et du Citoyen, 3 vols., Paris, 1908.
[29] Robinson v. Mandell, 20 Fed. Cas. 1027 (No. 11,959) (C. C. D. Mass. 1868).
[30] Swain v. Alabama, 380 U. S. 202 (1965).
[31] L. H. Tribe, Trial by mathematics: precision and ritual in the legal process, Harvard Law Rev., 84 (1971) 1329–1391.
[32] United States v. Coffin, Crim. No. 68-1 F (D. Mass., Aug. 14, 1968), *rev'd sub nom.* United States v. Spock, 416 F. 2d 165 (1st Cir. 1969).
[33] Whitus v. Georgia, 385 U. S. 545 (1967).
[34] H. Zeisel, Dr. Spock and the vanishing women jurors, Univ. Chicago Law Rev., 32 (1969) 1–18.
[35] H. Zeisel, And then there were none: the diminution of the federal jury, Univ. of Chicago Law Rev., 38 (1971) 710–724.

Incommensurability Proofs: A Pattern that Peters Out

The infinite descent argument works for some polygons—but not many.

E. J. Barbeau
University of Toronto
Toronto, Ontario, Canada M5S 1A1

The concept of incommensurable magnitudes can be traced back to the ancient Greeks. Recall that two magnitudes are incommensurable if their ratio is not given by a pair of positive integers. Our main interest here is not to make a historical study, but rather to examine the scope of a proof technique, which, though sometimes attributed to the Greeks, may well be of more recent origin. It is a method of infinite descent which can be applied to show the incommensurability of the diagonal and side of a square [**2**, p. 270]; [**11**, p. 44]; [**13**, p. 23].

This technique can also be employed to prove incommensurable the diagonal and a side of a regular pentagon and similarly the shortest diagonal and a side of a regular hexagon. Thus we have a method which works for three cases in a row, and we may well believe that it will apply to n-sided regular polygons for $n > 6$. Alas, such a hope is unfounded. While the shortest diagonal and a side of a regular n-gon may be incommensurable, the pattern that seems to emerge in establishing the result for $n = 4, 5, 6$ fails for $n > 6$. In this paper, we first discuss the reason for this failure. In doing so, we utilize some elementary algebraic facts about the roots of unity. Later, we consider the question of when the second shortest diagonal and a side of a regular polygon are incommensurable. Finally, we note the connection of these questions with the Euclidean algorithm.

The ratio of the shortest diagonal and a side of a regular n-gon

The argument to show that the diagonal and a side of a square are incommensurable goes like this. Let $ABCD$ be a square with diagonal AC (FIGURE 1). Choose points F and E on AC and BC respectively for which $AF = AB$ and $FE \perp AC$. Determine G to complete the square $CFEG$. This smaller square is related to the larger square in a special way. To see this, note that $BE = EF = FC$, and that $CF = AC - AB$. Substituting into the equation $CE = BC - BE$ gives $CE = 2AB - AC$. Let d and s be the diagonal and side, respectively, of the larger square and d_1 and s_1 those of the smaller. Our equations above show

$$d_1 = 2s - d \qquad s_1 = d - s.$$

The geometric procedure which produces from the given square $ABCD$ the smaller square $CFEG$ can be applied to the latter square to produce an even smaller square, and so on indefinitely. Now suppose that AC and BC are commensurable. Then there must exist a common measure x of which each is an integer multiple, so that there are positive integers m and n for which $d = nx$ and $s = mx$. Therefore $d_1 = n_1 x$ and $s_1 = m_1 x$, where $n_1 = 2m - n$ and $m_1 = n - m$ are both positive integers smaller than n and m respectively. If we construct the descending sequence of squares, we find that their respective diagonals and sides are $n_i x$ and $m_i x$ where n_i and m_i are positive integers for which

$$n > n_1 > n_2 \cdots > n_i > \cdots \geqslant 1, m > m_1 > m_2 \cdots > m_i > \cdots \geqslant 1.$$

But this puts an upper bound (of m) on the length of the sequence of squares that can be constructed, which contradicts our geometric conclusion. Thus, AC and AB must be incommensurable.

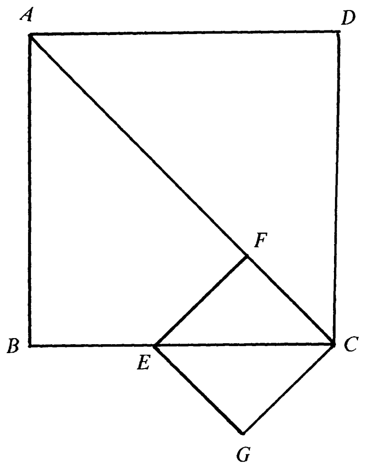


FIGURE 1

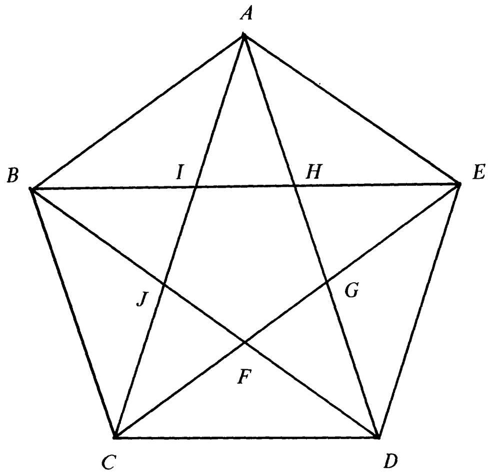


FIGURE 2

Is the same type of argument feasible for the shortest diagonal d and side s of other regular polygons? If so, the first step must be to construct from the larger polygon a smaller one with corresponding diagonal d_1 and side s_1 for which there are linear relations of the form

$$d_1 = ud + vs \qquad s_1 = wd + ys$$

with suitable integers u, v, w, y. This requires some ingenuity, for it can hardly be expected that every smaller figure that can be devised will lead to the right sort of relation, or that the equations, if they exist, can easily be found.

For the regular pentagon $ABCDE$ (FIGURE 2), a natural candidate for the smaller figure is the pentagon $FGHIJ$ framed by the diagonals of the larger. This case is treated by von Fritz [**15**, pp. 257–259] and Knorr [**12**, p. 30], and may have been the first occasion on which the incommensurability phenomenon was detected by the Greeks. Since the angles of the triangles ABJ, AIH, and CBI are 36°, 72°, 72° while those of the triangles ABC, IHG, and AIG are 108°, 36°, 36°, it follows that $AJ = AB = BC = IC$ and $AI = IG$. Hence

$$IJ = AJ + IC - AC = 2AB - AC$$

and

$$IG = AI = AC - IC = AC - AB.$$

If we denote, as before, d and s the length of a diagonal and a side, respectively, of the large pentagon and d_1 and s_1 those of the smaller pentagon $FGHIJ$, our equations show that in this case, $d_1 = d - s$ and $s_1 = 2s - d$. From these equations a proof of the incommensurability of d and s can be completed as for the square.

This method of obtaining a smaller pentagon from the larger differs sufficiently from that used for the square so that we do not have much guidance on how to proceed for more general situations. Accordingly, we will redo the pentagon case, and attempt to choose the smaller figure so that its side, rather than diagonal, is the difference of the original diagonal and side. Now, refer to FIGURE 3 in which the given pentagon is $ABCDE$, and diagonals BD, CE are drawn. Since the angle CFD is 108°, we make F a vertex of the smaller pentagon to be constructed. Produce BC to K and ED to L so that $CK = CF = FD = DL$. Then $DFCKL$ is a regular pentagon whose diagonal d_1 and side s_1 are related to those of the original by $d_1 = s$ and $s_1 = d - s$.

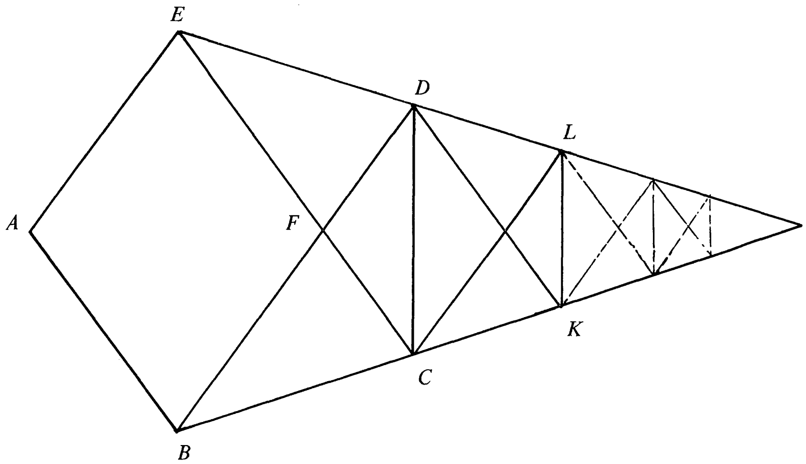


FIGURE 3

Our success with the pentagon case encourages us to try to show for the regular hexagon $ABCDEF$ (FIGURE 4), that diagonal AC and side AB are incommensurable. From AC, cut off $AG = AB$. With the aim of making GC the side of a smaller hexagon, choose H on the line through BC so that angle AGH is 60° (the external angle of a regular hexagon). Since angle AGB equals 75°, B evidently lies between H and C. Thus, C, G, H are three vertices of a regular hexagon, one of whose shortest diagonals is CH. To determine how CH is related to AB and AC, extend FA to meet the line through CB at K. Then AKB is an equilateral triangle, and angles AKH and AGH are both 60°. Since AGK is isosceles, so is triangle KHG, whereupon $KH = GH = GC$. Also $KB = AB = BC$. Therefore

$$CH = KB + BC - KH = 2BC - (AC - AG) = 3BC - AC.$$

Thus the shortest diagonal d and side s of the large hexagon are related to the corresponding magnitudes d_1 and s_1 of the smaller constructed hexagon by the equations

$$d_1 = 3s - d \qquad s_1 = d - s.$$

Once again we can begin our infinite descent to prove that d and s are incommensurable.

For three successive values of n, we have constructed from a regular n-gon a smaller version whose side satisfies an equation of the type $s_1 = d - s$. Had the Greeks pursued this line of enquiry, they would have discovered for the cases $n > 6$ unsolvable problems as compelling and frustrating as angle-trisection or circle-squaring. To see why, it is necessary to recall some background material on roots of unity.

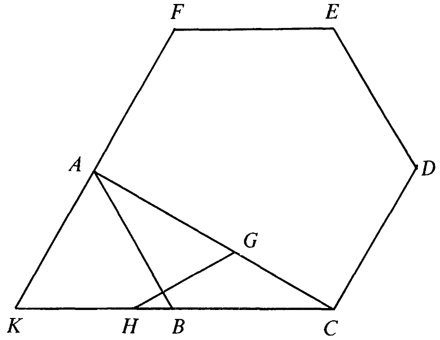


FIGURE 4

The reason for failure

We begin with some necessary algebra background; details of the theory can be found in many modern algebra textbooks, such as [**14**]. Suppose m is a positive integer. A complex number z for which $z^m = 1$ is an ***m*th root of unity**; if no smaller power of z is equal to 1, then z is a **primitive** mth root of unity and must be one of the numbers ζ^d where

$$\zeta = e^{2\pi i/m}$$

and d is relatively prime to m. In all, there are $\phi(m)$ distinct primitive mth roots of unity, where $\phi(m)$ is Euler's totient function (which counts the number of positive integers less than m which are relatively prime to m). It is well known that the function ϕ is multiplicative, so that if the prime factorization of m is $m = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, then $\phi(m) = \Pi_{1 \leqslant i \leqslant k} \phi(p_i^{\alpha_i})$. Since a number is relatively prime to $p_i^{\alpha_i}$ unless it is divisible by p_i, it follows that $\phi(p_i^{\alpha_i}) = p^{\alpha_i} - p^{\alpha_i - 1}$. Thus

$$\phi(m) = \prod_{1 \leqslant i \leqslant k} p_i^{\alpha_i - 1}(p_i - 1). \tag{1}$$

We will be interested in polynomials with integer coefficients which have ζ as a root; here $\phi(m)$ is of interest since it is the minimum degree of such a polynomial.

We will want to know all the values of m for which $\phi(m) \leqslant 4$. Except for $m = 1, 2$ (when $\phi(m) = 1$), $\phi(m)$ must be even. If p divides m, then (1) shows that $p - 1$ divides $\phi(m)$, so that when $\phi(m) \leqslant 4$, the only possible prime divisors of m are 2, 3, and 5. It is straightforward to determine that $\phi(m) = 2$ only when $m = 3$, 4, 6 and that $\phi(m) = 4$ only when $m = 5$, 8, 10, 12.

With this background, we are in a position to examine the range of validity of the proof technique discussed in our previous section. Suppose that $ABC\ldots$ is a regular n-gon with sides AB, $BC, \ldots$ of length s and shortest diagonals $AC, \ldots$ of length d (FIGURE 5). Suppose, also, that we have succeeded in constructing a smaller regular n-gon with side and shortest diagonal s_1 and d_1, respectively, for which $d_1 = ud + vs$ and $s_1 = d - s$, where u and v are integers. Let λ be the common value of d/s and d_1/s_1. Then

$$\lambda = (ud + vs)/(d - s) = (u\lambda + v)/(\lambda - 1)$$

whence

$$\lambda^2 - (u + 1)\lambda - v = 0. \tag{2}$$

If we denote angle ACB as θ, then FIGURE 5 shows that $\theta = \pi/n$ and $\lambda = 2\cos\theta = e^{i\theta} + e^{-i\theta}$. Our aim is to use equation (2) to find a polynomial of the fourth degree which has $e^{i\theta}$ as a root. There are two ways to proceed. One is to simply substitute $\lambda = e^{i\theta} + e^{-i\theta}$ in equation (2) and multiply the resulting equation by $e^{2i\theta}$. Alternatively, observe that $e^{i\theta}$, along with its complex conjugate, is a root of the polynomial $t^2 - \lambda t + 1$, and therefore a root of

$$\begin{aligned}(t^2 + 1 - \lambda t)\left[t^2 + 1 + (\lambda - (u + 1))t\right] &= (t^2 + 1)^2 - (u + 1)(t^2 + 1)t - \left[\lambda^2 - (u + 1)\lambda\right]t^2 \\ &= t^4 - (u + 1)t^3 + (2 - v)t^2 - (u + 1)t + 1. \qquad (3)\end{aligned}$$

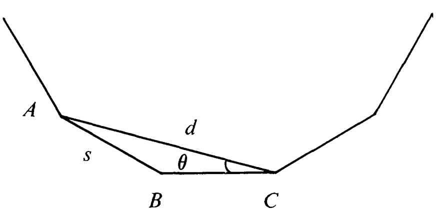


FIGURE 5

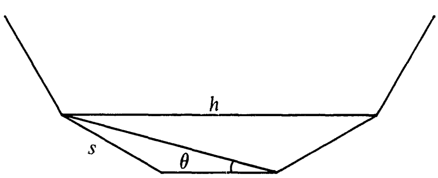


FIGURE 6

Since $\theta = \pi/n$, $e^{i\theta}$ is a primitive $2n$th root of unity. Since (3) is a polynomial with integer coefficients having $e^{i\theta}$ as a root, the degree $\phi(2n)$ of the minimum polynomial of $e^{i\theta}$ cannot exceed 4. Since $n = 2$ is not geometrically meaningful and $n = 3$ is degenerate, this leaves $n = 4$, 5, 6 as the only possibilities. Thus our assumption that we could write $s_1 = d - s$ is only feasible for the cases already proved. Even if we try to get $s_1 = wd + ys$ with other integers w, y, essentially the same argument restricts us to the same values of n. For $n > 6$, the method of proof of incommensurability of shortest diagonal and side peters out.

Other diagonals

The next natural problem on which to try the method of proof is to compare the length of a side to the second shortest diagonal of a regular n-gon. The smallest value of n for which a regular n-gon has diagonals of more than one length is 6, and this case is easily disposed of. The diagonal in question passes through the centre of the hexagon and is equal in length to twice the side. Thus they are commensurable. Now let $n > 6$, $\theta = \pi/n$ (as before), h be the length of the second shortest diagonal, s be the side length and $\mu = h/s$ (FIGURE 6). It is easily checked that $1 + 2\cos 2\theta = \mu$. If it were possible to construct a similar n-gon with second shortest diagonal and side respectively equal to h_1 and s_1 such that, for integers u, v, w, y, we have that $h_1 = uh + vs$ and $s_1 = wh + ys$, then, as we saw before, μ must be a root of quadratic polynomial. The primitive nth root of unity, $e^{2i\theta}$, satisfies

$$e^{2i\theta} + e^{-2i\theta} = \mu - 1$$

and so can be shown to be a root of a polynomial of degree 4 with integer coefficients. Hence $\phi(n)$ cannot exceed 4, so that n must be 8, 10, or 12. But can we actually find the required construction for each of these cases?

In FIGURE 7, let $ABCD\ldots$ be a regular n-gon, with h and s the lengths of AD and BC respectively. Since $n > 6$, $AD > 2BC$. Determine E on AD so that $AE = 2BC$. The smaller n-gon sought will have as two of its vertices D and E. We consider the cases $n = 8$, 10, and 12 individually.

(a) $n = 8$. Here $\theta = 22\frac{1}{2}°$, and it can be shown that the angles ACD and DEC are each equal to 5θ. Thus, the triangles ACD and DEC are similar, with the result that D, E and C are vertices of a regular octagon with EC as a shortest diagonal and CD as a second shortest diagonal. If h_1 and s_1 are the respective lengths of CD and DE, then $h_1 = s$ and $s_1 = h - 2s$.

(b) $n = 10$. In this case, $\theta = 18°$, angle DEC is equal to 108°, and each of the angles EDC and ECD is equal to 36°. Thus $DE = EC$. Produce DC to F and make $EC = CF$. Angle FEC equals 18°, and angle FED equals 126°, the same as angle ACD. Thus, triangles DEF and DCA are similar, and D, E, F are three vertices of a regular decagon whose side s_1 and second shortest diagonal h_1 are given by $s_1 = h - 2s$ and $h_1 = s + s_1 = h - s$.

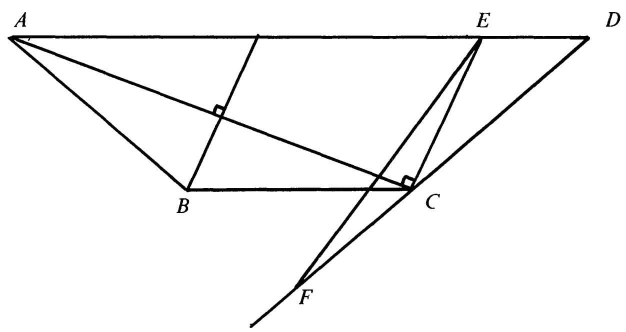


FIGURE 7

(c) $n = 12$. In this case, $\theta = 15°$. Determine F on DC produced so that angle DEF equals angle ACD, which is 135°. Then D, E, F are three vertices of a regular dodecagon and h_1 and s_1 are the respective lengths of DF and DE. By the law of sines, $DC : EC = \sin 105° : \sin 30°$ and $FC : EC = \sin 30° : \sin 15°$, from which $FC = CD$. (Is there a simple synthetic proof of this?) Therefore, $h_1 = 2s$ and $s_1 = h - 2s$.

In all three cases, the stage is set for an infinite descent argument which will contradict the hypothesis that the second shortest diagonal and side of the regular n-gon are commensurable.

When $n \geqslant 8$, a regular n-gon has diagonals of at least three distinct lengths, and we could try for a similar incommensurability proof for the third shortest diagonal and a side. If this were possible, then ν, the ratio of the length of this diagonal to the length of the side, would have to be a root of a quadratic polynomial with integer coefficients. Since

$$\nu = 2(\cos\theta + \cos 3\theta) = e^{-3i\theta} + e^{-i\theta} + e^{i\theta} + e^{3i\theta},$$

$e^{i\theta}$ must be the root of a polynomial of degree at most 12. This forces the degree, $\phi(2n)$, of the minimal polynomial of the primitive $2n$th root $e^{i\theta}$ not to exceed 12, so that n must be one of the integers 8, 9, 10, 11, 12, 13, 14, 15, 18, 21. It is a nice exercise in Galois theory to show that in none of these cases can ν be a quadratic irrational. The basic idea is to show that the orbit of ν under some automorphism of the field $\mathbf{Q}(e^{i\theta})$ over $\mathbf{Q}$ has more than two elements. As a consequence, the incommensurability of the third shortest diagonal and side of a regular n-gon cannot be established by an infinite descent argument.

The fact that the infinite descent argument is not possible for all but a few n-gons does not mean that other types of argument are not available. For example, in the shortest diagonal case it is straightforward to establish the irrationality of $\cos(\pi/n)$ for $n \geqslant 4$ [**1**, p. 28, number 157] and hence the incommensurability of d and s (recall $\lambda = d/s = 2\cos(\pi/n)$). For the case of the kth shortest diagonal, an analysis of the quantity

$$\zeta^{-k} + \zeta^{-k+2} + \zeta^{-k+4} + \cdots + \zeta^{k-4} + \zeta^{k-2} + \zeta^{k},$$

with ζ a primitive root of unity, is required. It would be nice to have a general argument to determine its algebraic degree over the rationals.

Euclid's criterion for incommensurability

Proposition 2 of Book X of Euclid's *Elements* gives a test for the incommensurability of two magnitudes:

> *If, when the less of two unequal magnitudes is continually subtracted in turn from the greater, that which is left never measures the one before it, the magnitudes will be incommensurable.* [**4**]

The practical difficulty with this algorithm is in establishing when the process actually fails to terminate. In some cases, this is achieved by demonstrating that it cycles. Let us see what happens when we apply this Euclidean algorithm or **anthyphairesis** ("taking away or stealing in turn") to the examples discussed earlier.

With the same notation as before (and with the obvious meaning for higher-order subscripts), the problem for the square is to show that d and s are incommensurable. Following Euclid's proposition, we take the lesser of the pair (d, s) from the greater and get the nonzero remainder $s_1 = d - s$. Now "that which is left" is s_1 and "the one before it" is s; taking the larger of the pair (s, s_1) from the smaller gives $d_1 = s - s_1$. The next pair to be examined is (s_1, d_1) with d_1 the larger member; the remainder this time is $s_2 = d_1 - s_1$. Thus, the anthyphairesis yields the successive equations

$$d = s + s_1, \quad s = s_1 + d_1, \quad d_1 = s_1 + s_2, \quad s_1 = s_2 + d_2,$$

and it is clear that we can continue this process indefinitely with nonzero remainders.

Similarly, for the pentagon case the algorithm does not terminate. For the first construction, we have

$$d=s+d_1, \quad s=d_1+s_1,$$

and so on, while for the second construction,

$$d=s+s_1=d_1+s_1, \quad d_1=d_2+s_2,$$

and so on. We skip the hexagon for a moment, and pass to the cases of the second shortest diagonal, where the algorithm starts with the pair (h,s). For the octagon, the first two steps give

$$h=s+(s+s_1)=h_1+(h_1+h_2)$$

and

$$(h_1+h_2)=h_1+h_2;$$

these are conveniently combined into a single step $h=2h_1+h_2$ (so that we take away from the larger magnitude the largest integer multiple of the smaller, as in the modern version of the Euclidean algorithm). The continuation is

$$h=2h_1+h_2, \quad h_1=2h_2+h_3, \quad h_2=2h_3+h_4,$$

and so on. Likewise, the algorithm applied to the decagon gives

$$h=2s+s_1, \quad s=s_1+(h_1-2s_1)=s_1+s_2, \quad s_1=s_2+s_3,$$

and so on. In both cases, we can continue the anthyphairesis forever.

The dodecagon case has a wrinkle. Our earlier computations give rise to the sequence of equations

$$h=2s+s_1=h_1+s_1, \quad h_1=2s_1+s_2=h_2+s_2, \text{ etc.}$$

But this does not quite correspond to Euclid's algorithm. Taking s from h leaves the remainder $s+s_1$; the next pair $(s,s+s_1)$ is responsible for the remainder s_1. Euclid's proposition specifies that we next deal with the pair (s,s_1). But, $s=s_1+\frac{1}{2}s_2$, so that our attempt to express the remainder in terms of quantities from our construction leads to fractions. To avoid the $\frac{1}{2}$, we compare $2s$ and s_1 rather than s and s_1.

A similar dodge is needed for the hexagon case. Since $d_1=3s-d$ and $s_1=d-s$, we get the following modified anthyphairesis, with a factor 2 inserted in every second step before the subtraction is made:

$$\begin{array}{ll} d=s+s_1 & \underline{2}s=s_1+d_1 \\ d_1=s_1+s_2 & \underline{2}s_1=s_2+d_2 \cdots . \end{array}$$

These equations can be used to find the standard continued fraction expansion of $\lambda=d/s$ for the hexagon case.

$$\lambda=\frac{d}{s}=\frac{s+s_1}{s}=1+\frac{s_1}{s}=1+\frac{2s_1}{2s}=1+\cfrac{2}{\cfrac{2s}{s_1}}=1+\cfrac{2}{\left(\cfrac{s_1+d_1}{s_1}\right)}$$

$$=1+\cfrac{2}{1+\cfrac{d_1}{s_1}}=1+\frac{2}{1+\lambda}. \tag{4}$$

Substituting $1+2/(1+\lambda)$ for λ on the right side of (4) gives

$$\lambda=1+\cfrac{2}{1+1+\cfrac{2}{1+\lambda}}=1+\cfrac{2}{2+\cfrac{2}{1+\lambda}}=1+\cfrac{1}{1+\cfrac{1}{1+\lambda}}. \tag{5}$$

Repeating the process with (5), we obtain

$$\lambda = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \left(1 + \cfrac{1}{1 + \cfrac{1}{1+\lambda}}\right)}} = 1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1+\lambda}}}} = \cdots .$$

Continuing on in this manner produces the continued fraction expansion for λ, which can be written compactly using the convention that in the representation below, each slash ($/$) is the bar of a fraction whose denominator is the entire expression following it:

$$\lambda = 1 + 1/1 + 1/2 + 1/1 + 1/2 + \cdots .$$

Similarly, a periodic continued fraction expansion can be found for the ratio d/s for the square and pentagon, and the ratio h/s for the octagon, decagon, and dodecagon. That the incommensurability of these ratios can be detected by a Euclidean algorithm which cycles is a reflection of the fact that the continued fraction of a quadratic irrational is periodic.

Our discussion here is not intended to reflect what was or might have been done by the Greeks. Although the question of the incommensurability of two magnitudes is generally believed to have been of major importance for them, very few references can be found in early Greek writings. Aristotle transmits the well-known argument for the incommensurability of diagonal and side of a square which shows the impossibility of $y^2 = 2x^2$ for a pair (x, y) of even integers, an argument which appears as Book X, Proposition 117, of Euclid's *Elements* (believed to be a later interpolation) [**4**]. Plato, in the *Theaetetus*, refers to the establishment by Theodorus of what is equivalent to the irrationality of the square roots of the nonsquare integers up to 17, but does not describe the method. Knorr [**12**, chap. III, IV, VI] discusses possible reconstructions in considerable detail. A recent investigator of Greek techniques, D. H. Fowler, argues that the Greeks were more familiar with and less disturbed by the simple fact of incommensurability than popular doctrine allows. He even suggests that the Greeks might have used anthyphairesis as a way of defining the ratio of two magnitudes. For his views and reconstructions, consult [**7**], [**8**], [**9**].

In this paper, I have preferred to present the incommensurability problem on its own, nonhistorical, merits. It is a good illustration to use in many undergraduate courses to show that what has been called the "strong law of small numbers" [**10**] can not always be relied on. Even seasoned mathematicians are tempted to believe that a method of proof which establishes a fact for $n = 4, 5, 6$ might indeed prove universal rather than abruptly peter out.

I am indebted to various colleagues and the referees for their comments on earlier drafts of this paper. In particular, I wish to thank D. H. Fowler for extensive remarks on Greek mathematics and for numerous preprints of his articles. The second argument I give for the pentagon is not original, but I cannot recall its source. Perhaps some reader can supply a reference.

References

[1] E. Barbeau, M. Klamkin, W. Moser, eds., 1001 Problems in High School Mathematics, Book 3, Canadian Mathematical Society, 1978.

[2] George Chrystal, Algebra, Part I, 1st ed., Black, 1886; 2nd ed., 1889.

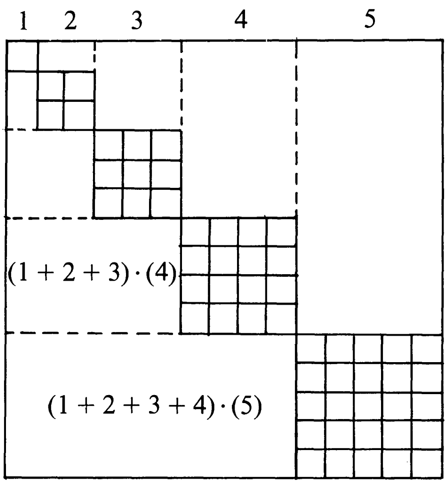
[3] H. S. M. Coxeter, Introduction to Geometry, Wiley, 1961.

[4] Euclid, Elements (translated with introduction and commentary by Sir T. L. Heath), 2nd ed., Dover, 1956.

[5] R. J. Evans and I. M. Isaacs, Fields generated by linear combinations of roots of unity II, Math. Scand., 43 (1978) 26–34.
[6] D. H. Fowler, Anthyphairetic ratio and Eudoxan proportion, Arch. Hist. Ex. Sci., 24 (1981) 69–71.
[7] ______, Book II of Euclid's Elements and a pre–Eudoxan theory of ratio, Arch. Hist. Ex. Sci., 22 (1980) 5–36.
[8] ______, Book II, of Euclid's Elements and a pre-Eudoxan theory of ratio, Part 2: Sides and diameters, Arch. Hist. Ex. Sci., (to appear).
[9] ______, Ratio in early Greek mathematics, Bull. Amer. Math. Soc., N. S., 1 (1979) 807–846.
[10] Martin Gardner, Mathematical Games: The strong law of small numbers, Sci. Amer. (6) 243 (December 1980) 18–28.
[11] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford, 1960.
[12] Wilbur R. Knorr, The Evolution of the Euclidean Elements, Reidel, 1975.
[13] Hans Rademacher and Otto Toeplitz, The Enjoyment of Mathematics, Princeton, 1957.
[14] B. L. van der Waerden, Modern Algebra, vol. 1, Ungar, New York, 1953.
[15] Kurt von Fritz, The discovery of incommensurability by Hippasus of Metapontum, Annals of Math., 46 (1945) 242–264.

Proof without words: Sum of squares

$$\sum_{k=1}^{n} k^2 = \left(\sum_{k=1}^{n} k\right)^2 - 2\sum_{k=1}^{n-1}\left[\left(\sum_{i=1}^{k} i\right)(k+1)\right].$$

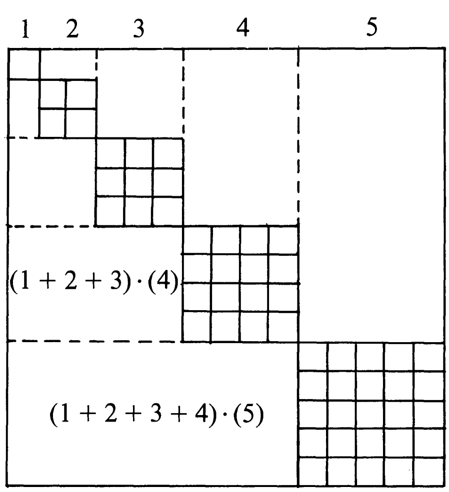


—JAMES O. CHILAKA
2626 Homecrest Ave.
Brooklyn, NY 11235

[5] R. J. Evans and I. M. Isaacs, Fields generated by linear combinations of roots of unity II, Math. Scand., 43 (1978) 26–34.
[6] D. H. Fowler, Anthyphairetic ratio and Eudoxan proportion, Arch. Hist. Ex. Sci., 24 (1981) 69–71.
[7] ______, Book II of Euclid's Elements and a pre-Eudoxan theory of ratio, Arch. Hist. Ex. Sci., 22 (1980) 5–36.
[8] ______, Book II, of Euclid's Elements and a pre-Eudoxan theory of ratio, Part 2: Sides and diameters, Arch. Hist. Ex. Sci., (to appear).
[9] ______, Ratio in early Greek mathematics, Bull. Amer. Math. Soc., N. S., 1 (1979) 807–846.
[10] Martin Gardner, Mathematical Games: The strong law of small numbers, Sci. Amer. (6) 243 (December 1980) 18–28.
[11] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford, 1960.
[12] Wilbur R. Knorr, The Evolution of the Euclidean Elements, Reidel, 1975.
[13] Hans Rademacher and Otto Toeplitz, The Enjoyment of Mathematics, Princeton, 1957.
[14] B. L. van der Waerden, Modern Algebra, vol. 1, Ungar, New York, 1953.
[15] Kurt von Fritz, The discovery of incommensurability by Hippasus of Metapontum, Annals of Math., 46 (1945) 242–264.

Proof without words:
Sum of squares

$$\sum_{k=1}^{n} k^2 = \left(\sum_{k=1}^{n} k\right)^2 - 2\sum_{k=1}^{n-1}\left[\left(\sum_{i=1}^{k} i\right)(k+1)\right].$$



—James O. Chilaka
2626 Homecrest Ave.
Brooklyn, NY 11235

Wroński's *Canons of Logarithms*

D. Bushaw
Washington State University
Pullman, WA 99164

As everyone knows, the electronic computer and the hand calculator have made tables of logarithms all but obsolete as practical aids to calculation. The history of these tables, however, glistens with examples of remarkable diligence and ingenuity. There may be no more striking example of cleverness in this area than Wroński's "canons of logarithms." This does not seem too strong a claim to make for a scheme that made it possible to put a seven-place table of common logarithms, in readable type, on a single page less than 17 cm by 22 cm in size.

J. M. Hoene-Wroński (1776–1853), born in Poland but resident for most of his life in France, was brilliant, erudite, industrious, versatile, and ambitious. His many projects ranged from the design of water works, navigational instruments, and railway car wheels through mathematics, astronomy, and other sciences to the farthest reaches of metaphysics—he devoted much of the last fifty years of his life to the elaboration and exposition of a vast philosophical system, called "messianism" in its later forms, somewhat similar to the systems of Fichte, Hegel, and other German idealist philosophers of the period. Wroński also had a difficult personality, and has been accused, not without plausibility, of arrogance, charlatanry, paranoia, and other blemishes of character. Thomas Muir called his style "exhaustingly wearisome." Reactions such as these may go far to account for the scant attention that his work, including the "canons of logarithms," has received.

Wroński died in poverty and relative obscurity, almost all of his voluminous writings forgotten except by a minute band of specialists and enthusiasts. The clearest exception is the "Wronskian determinant" (the term is due to Muir), well known to students of differential equations.

Wroński was apparently a tireless reckoner. One form this interest took was designing computational aids not only for scientists but also for nonscientists—school children, soldiers, business people, and others. Some of these devices should be called pre-electronic calculators, and in fact Wroński called one of his inventions the *calculateur universel* [**13**]. He left unfinished several projects of this kind, but the 1827 *Canons de logarithmes de H. W.* [**12**], with the companion leaflet *Prospectus des canons de logarithmes* [**11**], is a finished piece of work, although there are indications, both in these publications and in unpublished manuscripts, that he intended to develop the idea further. These two publications are now very rare, but a facsimile reprint of the first is included in a 1925 edition of Wroński's mathematical works.

Expositions of Wroński's system of canons have appeared in Russian [**1**]; Polish [**4**], [**5**], [**6**]; and Czech [**2**]. There seems to be no significant account of it, since Wroński's own, in any non-Slavic language. The present note is an attempt to do something about this. Taking the simplest of the six canons as typical of the set, we show how this canon is used, comment on why it works, and derive—for the first time—bounds for the errors, other than those due to rounding, associated with its use.

I	II	III	IV	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
Hoene Wroński's CANON OF LOGARITHMS No. 1 (1827)				000	414	792	139	461	761	041	304	553	788		A
				010	424	802	150	472	771	051	315	563	798		
				021	435	812	160	482	782	062	325	573	808		
				990	404	782	129	451	751	031	294	542	777		
				10. 0	11. 0	12. 0	13. 1	14. 1	15. 1	16. 2	17. 2	18. 2	19. 2		B
				20. 3	22. 3	24. 3	26. 4	28. 4	30. 4	32. 5	34. 5	36. 5	38. 5		
				40. 6	44. 6	48. 6	52. 7	56. 7	60. 7	64. 8	68. 8	72. 8	76. 8		
				50. 6	55. 7	60. 7	65. 8	70. 8	75. 8	80. 9	85. 9	90. 9	95. 9		
0.1	0.2	0.4	0.5	043	039	036	033	031	029	027	025	024	023	0.9	C
0.2	0.4	0.8	1.0	086	078	072	066	062	057	054	051	048	045	0.8	
0.3	0.6	1.2	1.5	128	117	107	099	092	086	081	076	072	068	0.7	
0.4	0.8	1.6	2.0	170	155	142	132	122	114	107	101	096	090	0.6	
0.5	1.0	2.0	2.5	212	193	177	164	152	142	134	126	119	113	0.5	
0.6	1.2	2.4	3.0	253	231	212	196	182	170	160	151	142	135	0.4	
0.7	1.4	2.8	3.5	294	268	246	228	212	198	186	175	166	157	0.3	
0.8	1.6	3.2	4.0	334	305	280	259	241	226	212	200	189	179	0.2	
0.9	1.8	3.6	4.5	374	342	314	291	271	253	238	224	212	201	0.1	
1.0	2.0	4.0	5.0	414	378	348	322	300	280	263	248	235	223	0.0	
.01	.02	.04	.05	04	04	04	03	03	03	03	03	02	02	02	D
.02	.04	.08	.10	09	08	07	07	06	06	05	05	05	05	04	
.03	.06	.12	.15	13	12	11	10	09	09	08	08	07	07	07	
.04	.08	.16	.20	17	16	14	13	12	12	11	10	10	09	09	
.05	.10	.20	.25	22	20	18	17	15	14	13	13	12	11	11	
.06	.12	.24	.30	26	24	22	20	19	17	16	15	14	14	13	
.07	.14	.28	.35	30	28	25	23	22	20	19	18	17	16	15	
.08	.16	.32	.40	35	31	29	27	25	23	22	20	19	18	17	
.09	.18	.36	.45	39	35	32	30	28	26	24	23	22	21	19	

FIGURE 1

One canon and how it is used

Wroński's 1827 booklet [**12**] includes instructions, examples, and theory for the use of six one-page tables of common logarithms. For good measure, it contains a summary of the "general solution of the fifth-degree equation." In the *Prospectus* Wroński argued that conventional logarithm tables are expensive, are bulky, have too many pages to turn, and frighten people. He had therefore devised his extremely compact "canons" (the term goes back to Napier) with none of those disadvantages. He may have been unaware that in this he was echoing Gauss, who in 1817 had written: "... the convenience of logarithmic tables intended for daily use decreases naturally in the same ratio as their bulk is increased" [**10**].

The six canons all operate in the same way, except for a small extra twist in the last and largest, which gives seven-place mantissas. Here "Canon No. 1" will be taken as typical of all six. A slightly modernized version of this table is shown above in FIGURE 1. The table is, in effect, a complete four-place table of common (base ten) logarithms!

The method for using the table will be illustrated with an example. The explanation may seem a bit long and involved, but with some practice one soons learns to perform the routine with speed and sureness.

Suppose the common logarithm of $n = 459.2$ is needed. We begin by moving the decimal point in n to get a number N satisfying $10 \leqslant N < 100$; so here $N = 45.92$. Next we look in section **B** of the table for the largest number not exceeding 45.92; this is 44, in the third row. It is called the "first part" of N. The first digit of the "first part" of the *mantissa* is then the 6 just to the right of the 44, and the remaining three digits of the mantissa, 435, appear in the corresponding position in section **A**. The whole "first part" of the mantissa of N is therefore .6435. (The rows and columns used in this example are marked with pointers in FIGURE 1.)

Next we look in column III of section **C** for the largest number that does not exceed the difference $45.92 - 44 = 1.92$. (Column III is used because the first parts of N and the mantissa were found in the *third* rows of sections **B** and **A**.) This number is 1.6, in the fourth row of section **C**, so 1.6 is the "middle part" of N. The "middle part" of the mantissa is then .0155, the last three digits of which appear where the same fourth row of section **C** crosses column (1); the digit immediately following the decimal point is always 0, and is omitted from the table to save space. For later use, note the number .6 in column (10) of the fourth row of section **C**.

The "final parts" of N and the mantissa are found in much the same way as the "middle parts," but by use of section **D**. The largest number in column III of section **D** that does not exceed $45.92 - 44 - 1.6 = .32$ is .32 itself, in the eighth row of section **D**. This is the "final part" of N, and the "final part" of the mantissa is .0029, found by looking in the same row and *the next column to the right* of the column (1) that was used before, namely column (2), where to save space the two 0's that immediately follow the decimal point are not shown.

To top off the calculation, we subtract the "final part" of the mantissa from the number just to its left in the table (remembering the invisible 0's) and multiply the difference by the .6 set aside earlier. This yields what Wroński called the "interpolation." In this example it is $(.0031 - .0029) \times .6 = .00012$.

(*Note*. Because N is exactly the sum of its first, middle, and final parts, the use of the table stops here. If this were not the case, one would again treat the residue as in the last two paragraphs, but with *all* numbers in section **D** multiplied by .1. If this still did not exhaust N, the same thing could be done again, but with all entries in section **D** multiplied by .01. However, by this time the additional parts of the mantissa would probably be negligible.)

Finally, one adds together the "parts" of the mantissa, including the "interpolation," and rounds the sum off to four places. The result is the mantissa for N. The characteristic for n is found in the usual way, and when combined with the mantissa gives the logarithm sought. In this example, the work can be summarized as follows:

	N	mantissa	
	45.92		
First parts	44.	.6435	
	1.92		
Middle parts	1.6	.0155	(store .6)
Final parts	.32	.0029	
Interpolation		.00012	
	Total	.66202	
	Rounded	.6620	
	log 459.2	2.6620.	

In fact, to six decimal places $\log 459.2 = 2.662002$.

Antilogarithms may be found similarly, moving from the larger "parts" to the smaller, but getting the "parts" of the number from those of the mantissa, not the reverse.

Why and how well the canon works

The method illustrated in the above example is based on the formula

$$\text{mant}\,(\omega x + \omega y + \omega z) \approx \text{mant}\,(\omega x) + \log\left(1 + \frac{y}{x}\right) + 10z\beta(x+1) + (1-y)[10z\beta(x) - 10z\beta(x+1)], \tag{1}$$

where "mant" means "mantissa of" and β is defined by

$$\beta(u) = \log\left(1 + \frac{1}{10u}\right) \qquad (u > 0). \tag{2}$$

The parameter ω takes the values 1, 2, 4, and 5, which correspond respectively to the four rows of section **B**, the four rows of section **A**, or the columns I, II, III, IV. The values of x are 10, 11,..., 19, as in the first row of section **B**; the values of y are 0 (which does not need to be shown) and .1, .2,..., 1.0 (in the intersection of column I with section **C**). The variable z can have any value satisfying $0 \leqslant z < .1$.

The first and middle parts of N are ωx and ωy. The quantity ωz includes the final part of N along with any residue of the type mentioned in the parenthetical note in the previous section. The successive steps based on section **D** deal with the successive decimal digits of z; this is legitimate because of the linear way in which z appears in formula (1).

The first part of the mantissa is the first term on the right side of (1), namely mant (ωx). Similarly, $\log(1+(y/x))$ is the middle part of the mantissa, $10z\beta(x+1)$ is the final part (plus possible accretions corresponding to the "residue" in N, if any) and the term $(1-y)[10z\beta(x)-10z\beta(x+1)]$, in which $1-y$ was represented by the stored factor .6 in the example, is the "interpolation." Except for the simple factor $1-y$, each term on the right side of (1) depends on only two of the variables, and this is what makes the two-dimensional layout of the canon possible.

Wroński obtained formula (1) by manipulating the results of several applications of Newton's interpolation formula (see for example [**9**], Chapter VI) and discarding terms he considered negligible. Instead of repeating this derivation, which is ingenious but not very instructive, let us turn to a more significant matter, namely the size of the errors incurred by using the right side of formula (1) to approximate the left side.

There is no evidence that Wroński dealt with this problem analytically. In particular, he is not likely to have used an explicit expression for the remainder in Newton's formula. An expression of this kind was first published in 1840, by Cauchy [**3**], and there is no reason to believe that Wroński anticipated Cauchy in this regard. Once he had discovered the general idea of the canons, he probably worked out the details more or less by trial and error, drawing on the excellent tables of logarithms already available at that time and on his own intuition and amazing energy.

Nevertheless, the errors can be estimated analytically. The conclusion is that for Canon No. 1 the error must lie between -9.275×10^{-6} and 5.375×10^{-6}, and these bounds are certainly good enough to justify a claim of four-place accuracy. But they are not good enough for "Canon No. 1 bis," where the same analysis applies but five-place accuracy is claimed. In fact, error analyses have been carried through for five of the six canons, and they show that the error bounds are consistent with the claims of accuracy in three cases and are almost good enough, but not quite, in the other two cases. Of course, there are also round-off errors. Wroński was well aware of them, and sensibly advised readers who thought a canon might give round-off errors excessive for their purposes to use a "higher" canon—one that gave more decimal places ([**12**], p. 28).

An error analysis for formula (1) (that is, for Canon No. 1 without round-off) appears in the next section. It is a somewhat unusual application of the techniques of elementary calculus in which convexity appears in a probably unexpected way.

Error analysis for Canon No. 1

If the right side of (1) is subtracted from the left, the resulting expression for the error E associated with (1) may be written

$$E(x,y,z)=\log\left(1+\frac{z}{x+y}\right)-10z[(1-y)\beta(x)+y\beta(x+1)]. \tag{3}$$

The problem is to find the maximum and minimum values of E for what will be called *admissible values* of x, y, and z:

$$x=10,\ 11,\ldots,\ 19;$$
$$0\leqslant y\leqslant 1;$$
$$0\leqslant z\leqslant .1.$$

(The variable y, like x, is really discrete, but it will be treated as continuous. Indeed, at one point in the analysis x too will be treated as continuous.)

We shall make several uses of the inequality

$$M\left(u-\frac{1}{2}u^2\right)\leqslant \log(1+u) \quad (u\geqslant 0;\ M=\log e=.43429\ldots), \tag{4}$$

which follows from Taylor's theorem; the inequality is sharp if $u>0$. Also, recall that a real-valued function f defined on an interval I is **convex** if for all $x_0, x_1 \in I$ and all t satisfying $0\leqslant t\leqslant 1$,

$$f((1-t)x_0+tx_1)\leqslant (1-t)f(x_0)+tf(x_1).$$

A well-known sufficient condition for convexity is that f'' should exist and be nowhere negative on the interval I.

With β as in (2), we set

$$B(x,y)=(1-y)\beta(x)+y\beta(x+1). \tag{5}$$

(The phrase "for all admissible values of the variables" should be understood here and at appropriate places in many of the statements that follow.)

From (2), $\beta''(u)>0$ for all $u>0$, so β is convex. Therefore,

$$\begin{aligned} B(x,y)&=(1-y)\beta(x)+y\beta(x+1)\\ &\geqslant \beta((1-y)x+y(x+1))=\beta(x+y). \end{aligned}$$

In short,

$$B(x,y)\geqslant \beta(x+y). \tag{6}$$

One other preliminary observation:

$$\frac{\partial^2}{\partial z^2}E(x,y,z)=-M(x+y+z)^{-2}<0. \tag{7}$$

Now we will find the minimum of E. From (6),

$$E(x,y,.1)=\beta(x+y)-B(x,y)\leqslant 0.$$

This fact, combined with (7) and the simple observation that $E(x,y,0)=0$, implies that

$$\phi(x,y)=E(x,y,.1)\leqslant E(x,y,z).$$

This reduces the problem of minimizing E to that of minimizing ϕ.

Consider the function γ defined by

$$\gamma(u)=M(10u^2+u)^{-1} \quad (u>0).$$

Then $\gamma''(u)=2M[u^{-3}-(u+.1)^{-3}]>0$, so γ is convex. From this in turn, with x treated as a continuous variable,

$$\frac{\partial}{\partial x}\phi(x,y)=(1-y)\gamma(x)+y\gamma(x+1)-\gamma(x+y)\geqslant 0.$$

This implies that $\rho(y)=\phi(10,y)\leqslant\phi(x,y)$, and the minimum of E will be that of ρ. Now

$$\begin{aligned} \rho'(y)&=-\gamma(10+y)+\beta(10)-\beta(11),\\ \rho''(y)&=M\left[(10+y)^{-2}-(10.1+y)^{-2}\right]>0. \end{aligned} \tag{8}$$

The quadratic equation obtained by setting ρ' equal to 0 has one admissible root, which to five places is .48814. Because of (8), ρ must have a minimum here. Thus the *minimum* of E is $E(10,.48814,.1)$, which to four significant figures is -9.275×10^{-6}.

We turn now to the problem of finding the maximum of E. For given admissible x and y, the one and only value of z for which $\partial E/\partial z$ vanishes is given by

$$z_c(x,y)=\frac{M}{10B(x,y)}-(x+y). \tag{9}$$

This value of z is admissible. In fact, from (9), (4), and (6),

$$z_c(x, y) \leqslant \frac{M}{10\beta(x+y)} - (x+y) < \frac{x+y}{20(x+y)-1}.$$

But $x+y \geqslant 10$, so from this display it follows that $z_c(x, y) < .1$. To complete the proof of

$$0 < z_c(x, y) < .1, \tag{10}$$

let

$$\lambda(x, y) = 10B(x, y)z_c(x, y) = M - 10(x+y)B(x, y).$$

For any fixed admissible x, the only critical value of y for λ is given by

$$y_c(x) = \frac{x\beta(x+1) - (x-1)\beta(x)}{2[\beta(x) - \beta(x+1)]}.$$

Numerical calculation shows that for each admissible value of x, $y_c(x)$ is an admissible value of y. Furthermore, as one easily verifies, $\partial^2\lambda/\partial y^2 > 0$. Thus for each x, λ has its minimum at $y_c(x)$. Additional calculation shows that $\lambda(x, y_c(x)) > 0$ for each x, so $\lambda(x, y) > 0$. The values of B are always positive, so from the definition of λ it follows that $z_c(x, y) > 0$.

From (7) it now follows that for fixed x and y, E attains its maximum at $z = z_c(x, y)$; that is, the maximum of E will be that of ψ, defined by

$$\psi(x, y) = E(x, y, z_c(x, y)). \tag{11}$$

From (3) and (9),

$$\psi(x, y) = \tau(10(x+y)B(x, y)), \tag{12}$$

where

$$\tau(u) = \log(M/u) - M + u \quad (u > 0).$$

The function τ satisfies

$$\tau'(u) = -M/u + 1 < 0 \quad \text{if } 0 < u < M. \tag{13}$$

To apply (13) we will need

$$0 < 100\beta(10) \leqslant 10(x+y)\beta(x+y) \leqslant 10(x+y)B(x, y) < M. \tag{14}$$

The last of these inequalities is equivalent to the left-hand part of (10). The third follows directly from (6). The second follows from $10 \leqslant x+y$ and the fact that for $\mu(u) = u\beta(u)$ and $u \geqslant 10$, by (4),

$$\mu'(u) = \beta(u) - \frac{M}{10u+1} > M\left(\frac{1}{10u} - \frac{1}{2(10u)^2}\right) - \frac{M}{10u+1} > 0.$$

Now, finally, by (11) – (14),

$$\begin{aligned}\psi(x, y) &= \tau(10(x+y)B(x, y)) \leqslant \tau(100\beta(10)) = \psi(10,0)\\ &= E(10,0, z_c(10,0)).\end{aligned}$$

Thus $\tau(100\beta(10))$ is the *maximum* of ψ, hence of E. To four significant figures, this maximum is 5.375×10^{-6}.

References

[1] A. Annenskii, Wroński's Tables of Logarithms, St. Petersburg, 1844 (Russian).
[2] J. Beneš, Hoene-Wroński's canons of logarithms, Časopis pro pěstovani mat. a fiz., 20 (1891) 24–42 (Czech).

[3] A. L. Cauchy, Sur les fonctions interpolaires, C. R. Acad. Sci. Paris, 11 (1840) 775–789. (Reprinted in Oeuvres, ser. 1, 5 (1885) 409–424.)
[4] Robert Ceceniowski, Hoene-Wroński and his tables of logarithms, Informator Kolejarza 1967 (Warsaw, 1966), pp. 165–174 (Polish).
[5] S. Dickstein, A Table of Logarithms by Hoene-Wroński, Wydawnictwo redakcyi "Prac matematyczno-fizycznych," Warsaw, 1890 (Polish).
[6] ______, Hoene-Wroński's Canons of Logarithms, Wydawnictwo redakcyi "Prac matematyczno-fizycznych," Warsaw, 1890 (Polish).
[7] ______, Sur les découvertes mathématiques de Wroński, Bibliotheca Mathematica, (2) 7 (1893) 9–14.
[8] ______, Hoene-Wroński: His Life and Works, Nakład Akademii Umiejętności, Warsaw, 1896 (Polish).
[9] Tomlinson Fort, Finite Differences and Difference Equations in the Real Domain, Oxford, 1948.
[10] C. F. Gauss, review of Abgekürzte logarithmisch-trigonometrische Tafeln,..., by Johann Pasquich, Göttingische gelehrte Anzeigen, Stück 158 (October 4, 1817) 1569–1574. (Reprinted in Werke, 3 (1876) 246–250, second printing.)
[11] [J.-M. Hoene-Wroński], Prospectus des canons de logarithmes, C. J. Trouve, n. p. [Paris?], n. d. [1827?].
[12] [J.-M. Hoene-Wroński], Canons de logarithmes de H. W., Jules Didot aîné, Paris, 1827. Facsimile with tables from [**6**] in Oeuvres mathématiques de Hoëné-Wroński, vol. 4, Paris, 1925.
[13] Hoëne-Wroński, Prospectus du calculateur universel et de ses diverses modifications, formant le calculateur nautique, le calculateur commercial, le calculateur de l'ingénieur, etc., etc., Beauce Rusand, Paris, 1823.

Apollonius by Inversion

A. Bruen
University of Western Ontario
London, Ontario, Canada N6A 5B9

J. C. Fisher
University of Regina
Regina, Saskatchewan, Canada S4S OA2

J. B. Wilker
University of Toronto
Toronto, Ontario, Canada M5S 1A1

How many circles can be drawn tangent to three given circles lying in the Euclidean plane? What additional possibilities are there if one or more of the circles are allowed to degenerate into a line or a point? These questions arose out of a problem posed and solved by Apollonius of Perga in the third century B. C.:

Given three objects, each of which may be a point, a line, or a circle, construct a circle which passes through each of the points and is tangent to the given lines and circles.

It is not known exactly how Apollonius solved this problem. Viète, working on hints coming from Pappus, proposed a restoration of the lost Apollonian construction. A later attempt at a

[3] A. L. Cauchy, Sur les fonctions interpolaires, C. R. Acad. Sci. Paris, 11 (1840) 775–789. (Reprinted in Oeuvres, ser. 1, 5 (1885) 409–424.)
[4] Robert Ceceniowski, Hoene-Wroński and his tables of logarithms, Informator Kolejarza 1967 (Warsaw, 1966), pp. 165–174 (Polish).
[5] S. Dickstein, A Table of Logarithms by Hoene-Wroński, Wydawnictwo redakcyi "Prac matematyczno-fizycznych," Warsaw, 1890 (Polish).
[6] ______, Hoene-Wroński's Canons of Logarithms, Wydawnictwo redakcyi "Prac matematyczno-fizycznych," Warsaw, 1890 (Polish).
[7] ______, Sur les découvertes mathématiques de Wroński, Bibliotheca Mathematica, (2) 7 (1893) 9–14.
[8] ______, Hoene-Wroński: His Life and Works, Nakład Akademii Umiejętności, Warsaw, 1896 (Polish).
[9] Tomlinson Fort, Finite Differences and Difference Equations in the Real Domain, Oxford, 1948.
[10] C. F. Gauss, review of Abgekürzte logarithmisch-trigonometrische Tafeln,..., by Johann Pasquich, Göttingische gelehrte Anzeigen, Stück 158 (October 4, 1817) 1569–1574. (Reprinted in Werke, 3 (1876) 246–250, second printing.)
[11] [J.–M. Hoene-Wroński], Prospectus des canons de logarithmes, C. J. Trouve, n. p. [Paris?], n. d. [1827?].
[12] [J.–M. Hoene-Wroński], Canons de logarithmes de H. W., Jules Didot aîné, Paris, 1827. Facsimile with tables from [**6**] in Oeuvres mathématiques de Hoëné-Wroński, vol. 4, Paris, 1925.
[13] Hoëne-Wroński, Prospectus du calculateur universel et de ses diverses modifications, formant le calculateur nautique, le calculateur commercial, le calculateur de l'ingénieur, etc., etc., Beauce Rusand, Paris, 1823.

Apollonius by Inversion

A. BRUEN
University of Western Ontario
London, Ontario, Canada N6A 5B9

J. C. FISHER
University of Regina
Regina, Saskatchewan, Canada S4S OA2

J. B. WILKER
University of Toronto
Toronto, Ontario, Canada M5S 1A1

How many circles can be drawn tangent to three given circles lying in the Euclidean plane? What additional possibilities are there if one or more of the circles are allowed to degenerate into a line or a point? These questions arose out of a problem posed and solved by Apollonius of Perga in the third century B. C.:

Given three objects, each of which may be a point, a line, or a circle, construct a circle which passes through each of the points and is tangent to the given lines and circles.

It is not known exactly how Apollonius solved this problem. Viète, working on hints coming from Pappus, proposed a restoration of the lost Apollonian construction. A later attempt at a restoration can be found in [**7**]. The various aspects of the problem have been a challenge to generations of geometers; Descartes, Newton, Euler, Gauss, and Cauchy are among the names occurring in its history. In the nineteenth century it served as a test case in the competition between rival schools of geometry. Elegant solutions eventually became common in textbooks (see, for example, [**8**] or [**10**]). Yet, according to N. A. Court in his historical survey [**1**], until the last quarter of the nineteenth century no one was particularly concerned with the number of solutions that the problem may have, which is the question that we wish to address.

In 1898 R. F. Muirhead [**9**] made the first noteworthy attempt at an exhaustive enumeration of the cases that can arise. We shall indicate here how his methods, although skillful, led to incomplete and unnecessarily complicated results. More recently, J. M. Fitz-Gerald [**6**] discussed the problem with special emphasis on the cases in which the three given objects are lines and circles with the property that no point is common to all three. In the present treatment we achieve an explicit listing of canonical forms for all the possible cases. Our work also serves as an independent proof of the main result in [**11**] to the effect that no specialization of the problem can produce exactly seven circles tangent to the given circles.

The classification of configurations of three given objects

The natural setting for an economical classification of cases for the problem is the inversive plane, which we shall think of as the Euclidean plane completed by a single point at infinity, P_∞. (See [**2**, pp. 77–95] or [**5**, pp. 103–131] for definitions and elementary theorems.) Since a line is simply an inversive circle which passes through P_∞, it is convenient to refer to all circles and lines as inversive circles, or ***i*-circles** for short. We use the word **object** as defined in the problem, that is, an object can be a line, circle, or point.

Since the incidence properties of objects are not altered by inversions, any enumeration of cases should be reduced to those which cannot be transformed into one another by a product of inversions. Muirhead failed to do this and therefore enumerated many more cases than necessary. In order to simplify our enumeration we use the following theorems of inversive geometry [**2**, § 6.5].

I. *A pair of disjoint i-circles can be inverted to a pair of concentric circles.*
II. *A pair of intersecting i-circles can be inverted to a pair of intersecting lines.*
III. *A pair of tangent i-circles can be inverted to a pair of parallel lines.*

In any case where two or more of these theorems apply we shall use the relevant theorem that appears first on our list. For instance, two intersecting circles disjoint from the third will be inverted to a pair of concentric circles and a circle intersecting none of them (rather than a pair of intersecting lines and a circle intersecting neither). This convention permits us to seek canonical forms in the following list of disjoint and exhaustive categories of three given objects.

1. Three points.
2. Two points and a line.
3. One point and two concentric circles.
4. One point and two intersecting lines.
5. One point and two parallel lines.
6. Two concentric circles and one *i*-circle.
7. Two intersecting lines and one *i*-circle meeting both lines.
8. Two parallel lines and one *i*-circle tangent to both lines.

For example, a triple of objects consisting of a point, line, and circle would be included under 3, 4, or 5 above depending on whether the line missed the circle, was secant to the circle, or was tangent to the circle. In our table of solutions to the problem of Apollonius, we make one exception to our stated convention: when the given point in 5 is P_∞, we picture this case as two circles tangent at the given point (so that this point appears in the picture).

It is *not* true that any two configurations (given triples of objects) with the same canonical form are inversively equivalent. This follows from the fact that the ratio of the radii of two concentric circles and the angle between two intersecting lines are inversive invariants [**3**]. However, the values of these parameters do not affect the *number* of tangent objects which are solutions to the problem for a given configuration, which is what we seek. This number depends only on the incidence and separation properties of the given objects. As our concluding remarks will indicate, it is an amusing exercise to determine this number by an inspection of the canonical form of the given configuration.

Sketches of the 33 possible canonical forms, 15 involving points and 18 involving only *i*-circles, are arranged in TABLE 1 and TABLE 2 respectively. The columns of these tables are labeled to indicate the number of solutions, where a solution is any object tangent to the three given objects. With Muirhead, we regard any object as being self-tangent. Thus, for example, a configuration of three mutually tangent *i*-circles with no point common to all three (category 8 in our list) gives rise to five solutions and not just two. In TABLES 1 and 2, the numbers under a diagram refer to Muirhead's classification [**9**, Table IV and Figures 50–114]. Note that his classification is incomplete as well as redundant.

We also find it convenient to include in TABLE 2 the descriptive labels of Fitz-Gerald who uses the symbol I for each intersecting pair of given *i*-circles, the symbol T for each tangent pair, and the symbol S for a given *i*-circle which separates the other two. As he points out [**6**, p. 18], it is unnecessary to specify which circles are involved since we are concerned only with the inversive relationship between the given *i*-circles taken in pairs and (should none of the circles intersect) whether or not one of the circles separates the other two. When the given *i*-circles have a point in common we put a bracket around the Fitz-Gerald label. Thus the appropriate label for three mutually tangent *i*-circles is either TTT (no point in common) or $[TTT]$ (a point in common). We note that almost all of the cases may be distinguished by their labels; the exceptions are III and $[III]$ and here we resort to subscripts.

We invite the reader to reproduce our 33 canonical forms of given configurations by systematically examining the eight categories listed above. As a typical example, here is how one might argue that category 7 leads to seven canonical forms. Begin with a pair of lines that meet in a point O (and, of course, in P_∞). The third given *i*-circle, which we'll call γ, can (a) intersect both lines, (b) intersect one line and be tangent to the other, or (c) be tangent to both. The *i*-circle γ cannot be disjoint from either line since that possibility belongs to category 6. One must now investigate each subcase in turn.

(a) The *i*-circle γ can intersect both lines in three distinct ways. First, γ can pass through the two points of intersection of the lines; this is $[III]_2$. Or, γ can pass through exactly one of the points of intersection; this is $[III]_1$. (Note that in TABLE 2 we show γ passing through P_∞ but we could equally well have shown γ passing through O.) Finally, γ can miss both O and P_∞; the two subcases are γ separates O from P_∞, which is III_2, or γ doesn't separate O from P_∞, which is III_1.

(b) The *i*-circle γ can be tangent to just one of the lines in two distinct ways. First, γ is tangent at P_∞ (or equivalently at O); this is $[IIT]$. Otherwise γ is tangent at some other point of one of the lines; this is IIT.

(c) There is essentially only one way for a circle to be tangent to both lines of an intersecting pair, namely ITT.

It is possible to adopt various other conventions for enumerating solutions; for example, one may argue that neither a point nor a given object should be regarded as a solution. However, it is clear from the tables that no convention would give rise to a case of seven tangent objects.

There appears to be some confusion in the literature as to what constitutes the "general case" of the problem. We may reasonably regard three *i*-circles to be in general position if they have *no common point* and if *no two are tangent*. The relevant cases are then $\varnothing$, III_1, III_2, II, I, and S; and our results are consistent with the statement attributed to Sturm [**6**], [**11**] that the number of circles tangent to three circles in general position is either 8, 4, or 0.

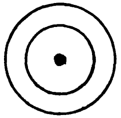
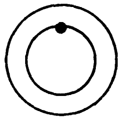
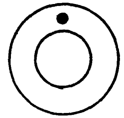
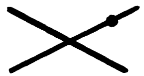
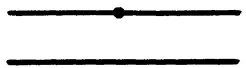
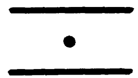
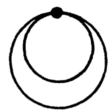
CONFIGURATION	NUMBER OF SOLUTIONS 0	1	2	3	4	∞
1. Three points.		 *				
2. Two points and a line.	 114, *	 #  #	 114, *			
3. One point and two concentric circles.	 50, 52, *		 #		 50, 52, *	
4. One point and two intersecting lines.		 #	 #  51, *			
5. One point and two parallel lines.		 #	 #	 #		 #

TABLE 1. **The fifteen canonical forms in which at least one of the given objects is a point. Numbers below a configuration refer to Figures 50–114 in Muirhead's classification [9]; an asterisk refers to his Table IV. The symbol # indicates that the configuration is not considered in [9].**

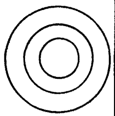
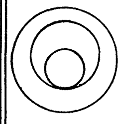
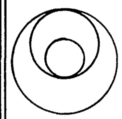
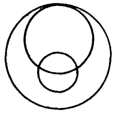
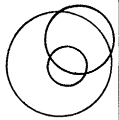
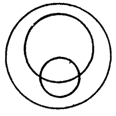
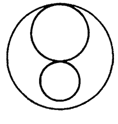
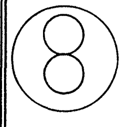
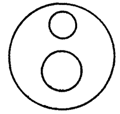
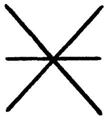
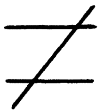
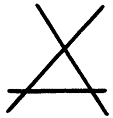
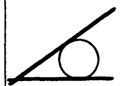
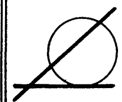
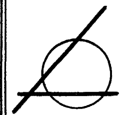
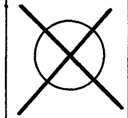
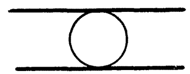
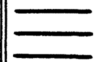
CONFIGURATION	NUMBER OF SOLUTIONS											
	0	2	3	4			5		6	8		∞
6. Two concentric circles and one *i*-circle.	 57, 58 103, 104 *S*	 71–74 *ST*	 92 *STT*	 76–81 *IT*	 63, 64 108–110 *II*	 59–62 105–107 *I*	 89, 91 *TT*		 69, 70, 75 *T*	 55, 56, 102 ∅		
7. Two intersecting lines and one *i*-circle meeting both.		 99 $[III]_2$	 97, 98 [*IIT*]				 86–88 $[III]_1$	 93–95 *ITT*	 82–85 *IIT*	 65–67 111, 113 III_1	 68, 112 III_2	
8. Two parallel lines and one *i*-circle tangent to both.							 # *TTT*					 100, 101 [*TTT*]

TABLE 2. **The eighteen canonical forms in which all three given objects are *i*-circles. Numbers below a configuration refer to Figures 50–114 in Muirhead's classification [9]; the symbol # indicates the configuration was not considered in [9]. The letter symbols under each configuration are adapted from Fitz-Gerald's descriptive labels [6].**

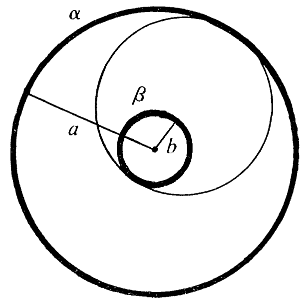


FIGURE 1. **The solution circles of radius $(a+b)/2$ for categories 3 and 6.**

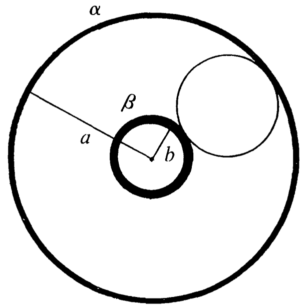


FIGURE 2. **The solution circles of radius $(a-b)/2$ for categories 3 and 6.**

Counting the solutions to the problem

When all possible configurations of three given objects have been enumerated, we can address the question of counting the number of objects tangent to the three given ones in each canonical arrangement. We shall indicate how one might investigate those cases in which a pair of the given objects are *i*-circles, say α and β.

H. S. M. Coxeter [**4**, section 5] has illustrated the process by a careful discussion of the two possibilities denoted by S and $\varnothing$ in TABLE 2, category 6; these arise when the three given objects are disjoint *i*-circles. Actually, his argument applies to all canonical forms that fall into our category 3 or 6. These are the cases in which α and β are concentric circles whose radii satisfy $a > b$. Since α and β are concentric, the circles that touch both consist of two families of congruent circles in the closed annulus bounded by α and β: one family having radius $(a+b)/2$ (FIGURE 1) and the other having radius $(a-b)/2$ (FIGURE 2). One can imagine a potential solution circle rolling about the annulus like a hoop until it reaches a position where it touches the third given object.

An analogous argument is valid for those canonical forms in which α and β are parallel lines (as in category 5 or 8): the solution circles (that are disjoint from P_∞) are then congruent circles touching both α and β.

Finally, α and β could be a pair of intersecting lines as in category 4 or 7. We conclude this discussion by giving a detailed analysis of the particular case *IIT* shown in FIGURE 3. In this case it is clear that any *i*-circle tangent to the intersecting lines α and β must be a circle in one of the quadrants A, B, C, or D into which they divide the plane. A general circle of this sort may be thought of as belonging to a family of circles whose members grow continuously from very small circles near the point O to very large circles far out in their quadrant. FIGURE 4 shows the six "stages of growth" at which the general circles of quadrants B, C, and D become tangent to circle γ. FIGURE 5 is the image of FIGURE 4 under inversion with centre P. It serves as a check on our count of tangent *i*-circles.

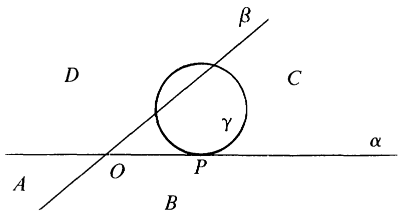


FIGURE 3. **The canonical form for the case *IIT*.**

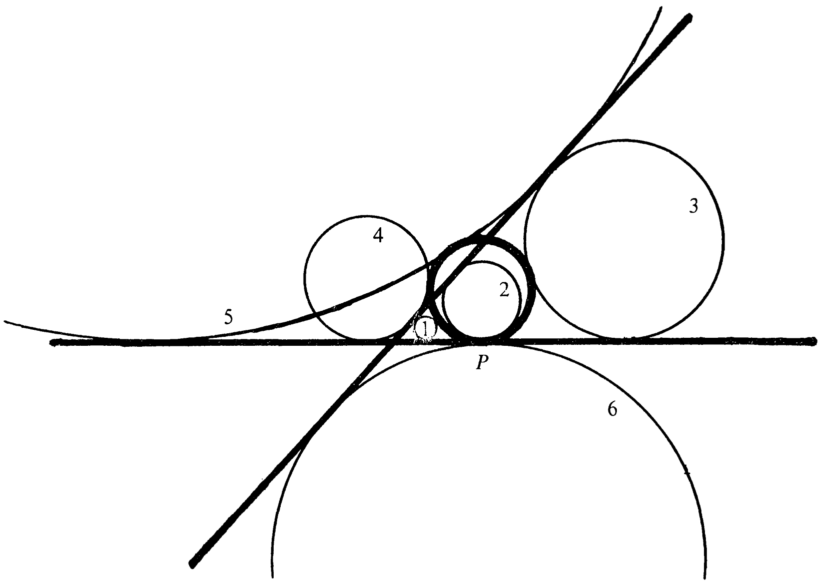


FIGURE 4. **Counting the solutions for the case *IIT*. The three given objects are drawn with heavy line.**

Once again we remind the reader that this paper calls for participation. Half the fun lies in reproducing the tables and this is a two-step project. First one has to list the 33 essentially different cases of the problem. Then in each of these cases one has to count solutions.

References

[1] N. A. Court, The problem of Apollonius, Math. Teacher, 54 (1961) 444–452.

[2] H. S. M. Coxeter, Introduction to Geometry, 2nd ed., Wiley, New York, 1969.

[3] ______, Inversive distance, Ann. Mat. Pura Appl., (4), 71 (1966) 73–88.

[4] ______, The problem of Apollonius, Amer. Math. Monthly, 75 (1968) 5–15. Reprinted in the Canad. Math. Bull., 11 (1968) 1–17.

[5] H. S. M. Coxeter and S. L. Greitzer, Geometry Revisited, New Math. Library, Mathematical Association of America, Washington, D.C., 1967.

[6] J. M. Fitz-Gerald, A note on the problem of Apollonius, J. Geom., 5 (1974) 15–26.

[7] Thomas Heath, A History of Greek Mathematics, vol. 2, Dover, New York, 1981.

[8] Roger A. Johnson, Advanced Euclidean Geometry (formerly, Modern Geometry), Dover, New York, 1960.

[9] R. F. Muirhead, On the number and nature of the solutions of the Apollonius contact problem, Proc. Edinburgh Math. Soc., 14 (1896) 135–147.

[10] D. Pedoe, Circles, A Mathematical View, Dover, New York, 1979.

[11] ______, The missing seventh circle, Elem. Math., 25 (1970) 14–15.

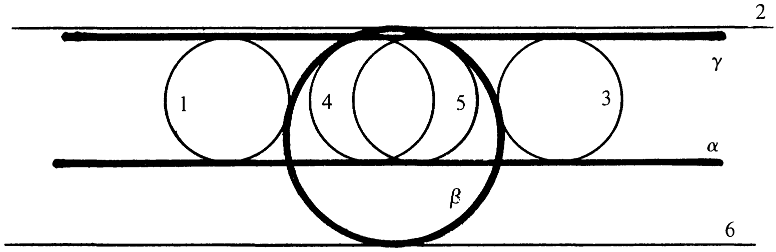


FIGURE 5. **Alternative form of the case *IIT*.**

Arc Length, Area, and the Arcsine Function

Andrew M. Rockett
Long Island University
Greenvale, NY 11548

Geometric interpretations of algebraic quantities provide an essential motivation for elementary calculus. The "calculus with analytic geometry" titles of many introductory texts allude to the reliance on pictorial images in the development of the derivative and the definite integral. While geometric considerations dominate most presentations of the trigonometric functions and their derivatives, there is a near universal switch to algebraic methods for the introduction and study of the inverse trigonometric functions (see [**6**], [**3**], and so on). This note shows how the geometric ideas used in the definitions of the definite integral, the trigonometric functions, and their derivatives may be continued in the discussion of the corresponding inverse functions. Reference will be made to the relation of these geometric ideas to the development of the theory of elliptic functions and to Euler's method of finding algebraic addition theorems for circular, hyperbolic, and lemniscate sines.

"Typical calculus" versus the arcsine as arc length

Taking [**6**] and [**3**] as our models, we note that in the typical modern textbook, after the definite integral has been defined, the applications include the area between two curves and the arc length formula. Since few integration techniques are available, the arc length problems are restricted to "nice" curves $y=f(x)$ such that the integral $\int_a^b\sqrt{1+f'(x)^2}\,dx$ is particularly simple and sometimes an author's apology is offered for the lack of interesting applications (see [**3**], p. 429).

The introduction of the trigonometric functions follows a review of radian measure as arc length measured from the point $(1,0)$ on the unit circle $x^2+y^2=1$. The sine and cosine of a real number θ are defined as the coordinates of the point (x,y) on the unit circle θ radians from $(1,0)$ (see Figure 1). Then the properties of $\sin\theta$ and $\cos\theta$ are derived from the symmetries of the circle and the other trigonometric functions are defined in terms of the sine and cosine. The derivatives of the sine and the cosine are found as consequences of $\lim_{\theta\to 0}(\sin\theta)/\theta=1$. This limit is established by equating arc length along the edge of the unit circle with the area of the sector determined by the arc (in Figure 2, $\theta=2\cdot$ area AOB) and then squeezing this area between two triangular regions.

After studying the calculus of the six trigonometric functions ("$f(x)$"), the corresponding inverse functions ("$f^{-1}(x)$") are sought by reversing the graphs ("$y=f(x)$" becomes "$x=f^{-1}(y)$"), making arbitrary choices for the "principal values" (see [**6**], pp. 295–6), and then calculating $D_x(f^{-1}(x))$ from the identity $f(f^{-1}(x))=x$.

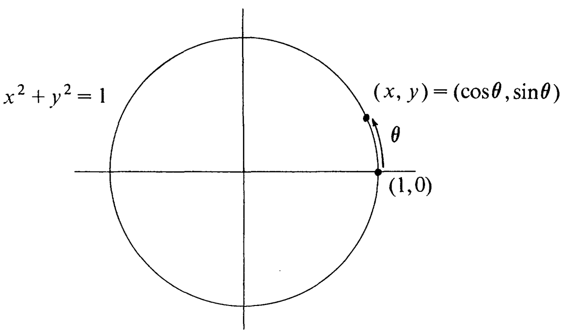


Figure 1

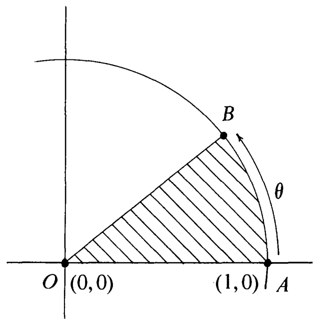


Figure 2

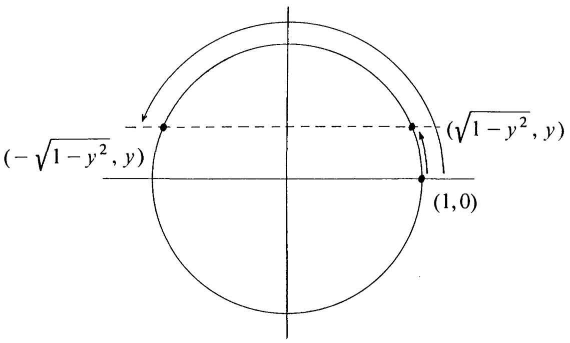


FIGURE 3

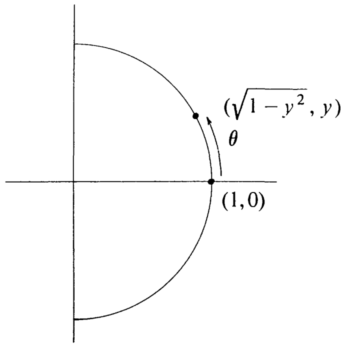


FIGURE 4

In contrast to exploiting the definitions and properties of inverse functions, the arcsine function can be approached in a more geometric way. Since $\sin\theta$ was defined as the y-coordinate of the point on the unit circle an arc length of θ away from $(1,0)$, a geometric attempt to invert this function would ask "given that the second coordinate of the point is y, from what arc length did it come?" There are two "small" answers to this question (see FIGURE 3) and then infinitely more separated from each other by multiples of 2π. The "principal value" of the inverse sine function may be introduced naturally as the smallest distance from the starting point, $(1,0)$, that will work, and this arc length may be called the "arc sine of y" and written "$\arcsin(y)$" or the "inverse sine of y" and written "$\sin^{-1}(y)$." We will use the "arcsine" notation for the remainder of our discussion. Thus the arcsine function has $\arcsin(y)=\theta$ where $-\pi/2\leqslant\theta\leqslant\pi/2$ and $\sin\theta=y$.

Since $\arcsin(y)$ is an arc length, the arc length formula $\int_a^b\sqrt{1+f'(t)^2}\,dt$ can be applied to $f(t)=\sqrt{1-t^2}$ from $t=0$ to $t=y$ (see FIGURE 4) to find that

$$\begin{aligned}\arcsin(y)&=\int_0^y\sqrt{1+f'(t)^2}\,dt\\&=\int_0^y\sqrt{1+\frac{t^2}{1-t^2}}\,dt\\&=\int_0^y\frac{1}{\sqrt{1-t^2}}\,dt\end{aligned}\tag{1}$$

and then, by the Fundamental Theorem of Calculus, it follows that

$$D_y(\arcsin(y))=\frac{1}{\sqrt{1-y^2}}.$$

Since $\arcsin(1)=\pi/2$, we also have a simple example of an improper integral:

$$\int_0^1\frac{1}{\sqrt{1-t^2}}\,dt=\lim_{y\to1}\int_0^y\frac{1}{\sqrt{1-t^2}}\,dt=\frac{\pi}{2}.$$

For the inverse tangent, a similar argument yields an arctangent function $\arctan(w)=\theta$ with principal value $-\pi/2<\theta<\pi/2$ such that $\arctan(w)$ is the distance along the unit circle from $(1,0)$ to the point (x,y) with $y/x=w$. Since $x=\sqrt{1-y^2}$, we can solve $y/x=w$ to find $y=w/\sqrt{1+w^2}$. From the arc length formula (1) we have

$$\arctan(w)=\int_0^{w/\sqrt{1+w^2}}\frac{1}{\sqrt{1-t^2}}\,dt.$$

Making the change of variable $t = u/\sqrt{1+u^2}$, so that $t = 0$ when $u = 0$ and $t = w/\sqrt{1+w^2}$ when $u = w$, we find that

$$\arctan(w) = \int_0^w \sqrt{1+u^2} \frac{1}{\sqrt{1+u^2}(1+u^2)} du$$
$$= \int_0^w \frac{1}{1+u^2} du,$$

and so

$$D_w(\arctan(w)) = \frac{1}{1+w^2}.$$

For the inverse secant of w (where $|w| \geqslant 1$), if we seek the smallest positive arc length from $(1,0)$ to the point with first coordinate $1/w$, we obtain an arcsecant function $\operatorname{arcsec}(w) = \theta$ with principal value $0 \leqslant \theta < \pi/2$ and $\pi/2 < \theta \leqslant \pi$. For $w > 1$, the arc length θ is $\pi/2 - \phi$ (see FIGURE 5) so we have

$$\operatorname{arcsec}(w) = \frac{\pi}{2} - \arcsin(1/w)$$

and so

$$D_w(\operatorname{arcsec}(w)) = -\frac{1}{\sqrt{1-(1/w)^2}} \cdot \frac{-1}{w^2}$$
$$= \frac{1}{w\sqrt{w^2-1}} \quad \text{for } w > 1.$$

For $w < -1$, we have by symmetry that

$$\operatorname{arcsec}(w) = \pi - \operatorname{arcsec}(-w)$$

and so

$$D_w(\operatorname{arcsec}(w)) = -D_w(\operatorname{arcsec}(-w))$$

and thus

$$D_w(\operatorname{arcsec}(w)) = \frac{1}{|w|\sqrt{w^2-1}} \quad \text{for } |w| > 1.$$

(We omit discussion of the formula

$$\operatorname{arcsec}(w) = \int_0^{\sqrt{w^2-1}/w} \frac{1}{\sqrt{1-t^2}} dt \quad \text{for } w \geqslant 1$$

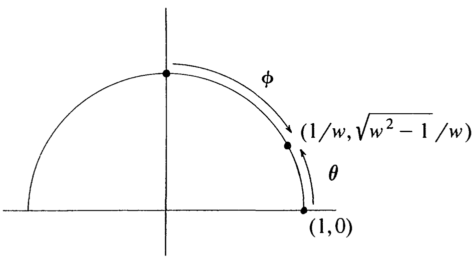


FIGURE 5

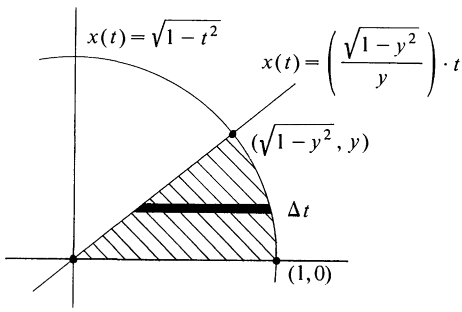


FIGURE 6

as it results in a subtle improper integral calculation.)

The formulas for the inverse cosine, cotangent, and cosecant can be obtained by similar arguments.

Area and the arcsine

Let us now study the arcsine as an area. Let $0 < y < 1$ be given and let A denote the area bounded by the line $x(t) = (\sqrt{1-y^2}/y) \cdot t$, the circle $x(t) = \sqrt{1-t^2}$, and the x-axis $t = 0$ (see FIGURE 6). Since arc length along the edge of the unit circle equals twice the area of the sector determined by the arc, we have that

$$\begin{aligned}\arcsin(y) &= 2 \cdot A \\ &= 2 \cdot \int_0^y \left(\sqrt{1-t^2} - \frac{\sqrt{1-y^2}}{y} \cdot t \right) dt \\ &= 2 \cdot \int_0^y \sqrt{1-t^2}\, dt - y\sqrt{1-y^2}. \end{aligned} \tag{2}$$

Differentiating (2), we have

$$\begin{aligned} D_y(\arcsin(y)) &= \sqrt{1-y^2} + \frac{y^2}{\sqrt{1-y^2}} \\ &= \frac{1}{\sqrt{1-y^2}}, \end{aligned}$$

as before.

Alternatively, our integral expression (2) for $\arcsin(y)$ provides a motivating example for the integration by parts formula $\int u \cdot dv = u \cdot v - \int v \cdot du$. Letting $u = \sqrt{1-t^2}$ and $dv = dt$, we have

$$\begin{aligned} \int \sqrt{1-t^2}\, dt &= t\sqrt{1-t^2} - \int t \frac{-t}{\sqrt{1-t^2}}\, dt \\ &= t\sqrt{1-t^2} - \int \frac{\left(\sqrt{1-t^2}\right)^2 - 1}{\sqrt{1-t^2}}\, dt \\ &= t\sqrt{1-t^2} - \int \sqrt{1-t^2}\, dt + \int \frac{1}{\sqrt{1-t^2}}\, dt \end{aligned}$$

so

$$2\int \sqrt{1-t^2}\, dt = t\sqrt{1-t^2} + \int \frac{1}{\sqrt{1-t^2}}\, dt.$$

But then

$$\begin{aligned} \arcsin(y) &= 2 \cdot \int_0^y \sqrt{1-t^2}\, dt - y\sqrt{1-y^2} \\ &= y\sqrt{1-y^2} + \int_0^y \frac{1}{\sqrt{1-t^2}}\, dt - y\sqrt{1-y^2} \\ &= \int_0^y \frac{1}{\sqrt{1-t^2}}\, dt, \end{aligned}$$

as before.

This development of the arcsine as the area bounded by the x-axis, the circle $x(t)=\sqrt{1-t^2}$ from $t=0$ to $t=y$, and the line from the origin to the point $(\sqrt{1-y^2}, y)$ may be adapted to the hyperbola $x(t)=\sqrt{1+t^2}$ to find the "inverse hyperbolic sine" $\sinh^{-1}(y)=\ln|\sqrt{1+y^2}+y|$. However, unlike the circle, there is no simple relation between arc length and area for the hyperbola (see [4], §6 and 7; for a representation of arc length on the hyperbola, see [1], §60, and the corresponding geometric construction in §61). On the other hand, a similar procedure to develop a "lemniscate sine" using the lemniscate instead of the circle must be carried out in terms of arc length (see [4], § 8).

Rationalization of the arcsine

We now turn to the problem of expressing the integrand in (1) as a rational function; that is, we wish to express $1-t^2$ as the square of a rational function. If, in the two-variable representation of Pythagorean triples,

$$(a^2-b^2)^2+(2ab)^2=(a^2+b^2)^2, \tag{3}$$

we let $a=1$ and $b=u$ then

$$(1-u^2)^2=(1+u^2)^2-(2u)^2$$

or

$$\left(\frac{1-u^2}{1+u^2}\right)^2=1-\left(\frac{2u}{1+u^2}\right)^2,$$

and so the substitution $t=2u/(1+u^2)$ will result in

$$\sqrt{1-t^2}=\frac{1-u^2}{1+u^2}$$

and

$$dt=\frac{(1+u^2)(2)-(2u)(2u)}{(1+u^2)^2}du=2\frac{1-u^2}{(1+u^2)^2}du.$$

Then the indefinite integral $\int 1/\sqrt{1-t^2}\,dt$ becomes

$$\int\frac{1+u^2}{1-u^2}\cdot 2\frac{1-u^2}{(1+u^2)^2}du=2\cdot\int\frac{1}{1+u^2}du$$

and the integrand has been rationalized. Solving $t=2u/(1+u^2)$ for u, we find that $u=(1\pm\sqrt{1-t^2})/t$. Equating t with $\sin\theta$, we see that $\sqrt{1-t^2}=\cos\theta$ and the expression for u with the minus sign becomes

$$u=(1-\cos\theta)/\sin\theta=\tan(\theta/2)$$

and we have derived the "$u=\tan(\theta/2)$" substitution for the rationalization of trigonometric integrals. (Compare this development with the typical "it has been discovered that..." treatment in [6], p. 368.)

Euler's sine sum formula

In the introductory section of [5], Siegel describes Fagnano's study of arc length on the lemniscate (which follows our development of the arcsine on the circle) and speculates that Fagnano's 1718 discovery of a geometric construction to double arc length on the lemniscate resulted from his attempt to rationalize the integrand of the lemniscate sine. Thirty-five years later, Euler extended Fagnano's doubling theorem to an algebraic addition theorem for the lemniscate sine and he shortly thereafter generalized his discovery to elliptic integrals. Siegel ([5], p. 10) describes the aim of his first chapter to be the fuller understanding of Euler's result from the viewpoint of analytic functions in their full domain of definition. We conclude our discussion of the arcsine with Euler's algebraic addition theorem for the arcsine adapted from § 585 and 586 of [**2**] (see also [**2**], Caput VI for Euler's study of elliptic integrals; [**4**], Chapter 4; and [**5**], § 2).

Let $-\pi/2<\tau<\pi/2$ be a fixed angle and let $-\pi/2<\theta,\phi<\pi/2$ be any two angles with $\theta+\phi=\tau$. Since the sum $\theta+\phi$ is constant, $d(\theta+\phi)=0$, and if we set $u=\sin\theta$ and $v=\sin\phi$, we can rewrite $d(\theta+\phi)=0$ as

$$d\left(\int_0^u\frac{dt}{\sqrt{1-t^2}}+\int_0^v\frac{dt}{\sqrt{1-t^2}}\right)=0$$

and so we have the differential equation

$$\frac{du}{\sqrt{1-u^2}}+\frac{dv}{\sqrt{1-v^2}}=0. \tag{4}$$

We seek an algebraic solution of (4) subject to the condition that $\theta+\phi=\tau$.

Euler's main observation was that if we begin with the symmetric second-order equation

$$u^2+2auv+v^2=K^2, \tag{5}$$

where a and K are constants, we can complete the square on the left side in either u or v. In the first case, we have

$$u^2+2auv+a^2v^2=K^2+(a^2-1)v^2,$$

so that

$$u+av=\sqrt{K^2+(a^2-1)v^2}, \tag{6}$$

while in the second, we have

$$v^2+2auv+a^2u^2=K^2+(a^2-1)u^2,$$

so that

$$v+au=\sqrt{K^2+(a^2-1)u^2}. \tag{7}$$

If we differentiate equation (5), we find

$$2udu+2avdu+2audv+2vdv=0,$$

which becomes, after collecting terms and using (6) and (7),

$$\frac{du}{\sqrt{K^2+(a^2-1)u^2}}+\frac{dv}{\sqrt{K^2+(a^2-1)v^2}}=0. \tag{8}$$

If we set $a^2-1=-K^2$ then equation (8) is the same as (4) and so $a=\sqrt{1-K^2}$. If we can solve (5) for K using this value of a, we will find an algebraic solution of (4) that expresses the constant value $\sin(\theta+\phi)$ in terms of $\sin\theta$ and $\sin\phi$.

Substituting our value for a into (5) and rearranging, we have

$$(u^2+v^2)-K^2=-2\sqrt{1-K^2}\,uv$$

and squaring both sides gives

$$(u^2+v^2)^2-2K^2(u^2+v^2)+K^4=4(1-K^2)u^2v^2. \tag{9}$$

By setting $a=u$ and $b=v$ in equation (3) and solving for $(u^2+v^2)^2$, we can rewrite (9) as

$$(u^2-v^2)^2-2K^2\big((u^2+v^2)-2u^2v^2\big)+K^4=0. \tag{10}$$

Since (10) is a quadratic in K^2, we can complete the square with respect to K^2 to obtain

$$\big(K^2+(2u^2v^2-(u^2+v^2))\big)^2=\big((u^2+v^2)-2u^2v^2\big)^2-(u^2-v^2)^2. \tag{11}$$

By expanding, collecting terms, and removing common factors, the right side of (11) may be rewritten as $4u^2v^2(u^2v^2 - (u^2+v^2)+1)$. Taking square roots, we find

$$K^2 = (u^2+v^2) - 2u^2v^2 + 2uv\sqrt{(1-u^2)(1-v^2)}\,. \tag{12}$$

Regrouping the expression $(u^2+v^2)-2u^2v^2$ as $(u^2-u^2v^2)+(v^2-v^2u^2)$, we see that the right side of (12) is a perfect square and so

$$K = u\sqrt{1-v^2} + v\sqrt{1-u^2}\,. \tag{13}$$

Changing back to angles, K is the constant $\sin(\theta+\phi)$ while $\sqrt{1-v^2} = \sqrt{1-\sin^2\phi} = \cos\phi$ and $\sqrt{1-u^2} = \cos\theta$ so we have the sine sum formula

$$\sin(\theta+\phi) = \sin\theta\cos\phi + \sin\phi\cos\theta\,. \tag{14}$$

Since the fixed value of τ played no part in our calculations, we have established (14) for any τ.

References

[1] Arthur Cayley, An Elementary Treatise on Elliptic Functions, 2nd ed., 1895. Reprinted by Dover Pub., New York, 1961.

[2] Leonhard Euler, Institutiones Calculi Integralis, Volumen Primum, 1768. Reprinted as Leonhardi Euleri Opera Omnia, Series I, Opera Mathematica, Volumen XI, Engel et Schlesinger, eds., B. G. Teubneri, Leipzig, 1913.

[3] Louis Leithold, The Calculus with Analytic Geometry, 4th ed., Harper & Row, New York, 1981.

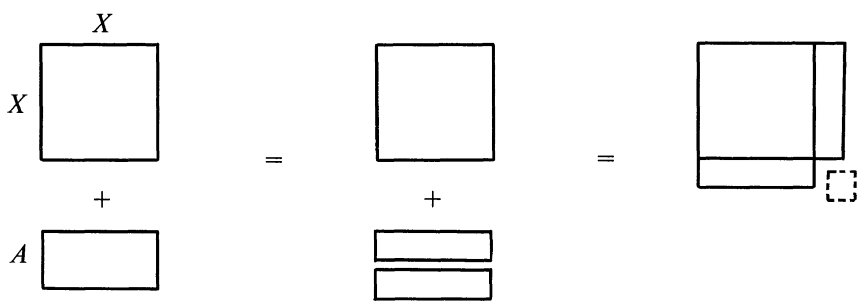
[4] A. I. Markushevich, The Remarkable Sine Functions (translated by Scripta Technica, Inc.), Amer. Elsevier, New York, 1966.

[5] C. L. Siegel, Topics in Complex Function Theory, vol. I: Elliptic Functions and Uniformization Theory (translated by Shenitzer and Solitar), Wiley–Interscience, New York, 1969.

[6] George B. Thomas and Ross L. Finney, Calculus and Analytic Geometry, 5th ed., Addison–Wesley, Reading, 1979.

Proof without words: Completing the square

$$X^2 + AX = (X + A/2)^2 - (A/2)^2$$



—CHARLES D. GALLANT
St. Francis Xavier University
Antigonish, Nova Scotia

By expanding, collecting terms, and removing common factors, the right side of (11) may be rewritten as $4u^2v^2(u^2v^2-(u^2+v^2)+1)$. Taking square roots, we find

$$K^2=(u^2+v^2)-2u^2v^2+2uv\sqrt{(1-u^2)(1-v^2)}\,. \tag{12}$$

Regrouping the expression $(u^2+v^2)-2u^2v^2$ as $(u^2-u^2v^2)+(v^2-v^2u^2)$, we see that the right side of (12) is a perfect square and so

$$K=u\sqrt{1-v^2}+v\sqrt{1-u^2}\,. \tag{13}$$

Changing back to angles, K is the constant $\sin(\theta+\phi)$ while $\sqrt{1-v^2}=\sqrt{1-\sin^2\phi}=\cos\phi$ and $\sqrt{1-u^2}=\cos\theta$ so we have the sine sum formula

$$\sin(\theta+\phi)=\sin\theta\cos\phi+\sin\phi\cos\theta. \tag{14}$$

Since the fixed value of τ played no part in our calculations, we have established (14) for any τ.

References

[1] Arthur Cayley, An Elementary Treatise on Elliptic Functions, 2nd ed., 1895. Reprinted by Dover Pub., New York, 1961.

[2] Leonhard Euler, Institutiones Calculi Integralis, Volumen Primum, 1768. Reprinted as Leonhardi Euleri Opera Omnia, Series I, Opera Mathematica, Volumen XI, Engel et Schlesinger, eds., B. G. Teubneri, Leipzig, 1913.

[3] Louis Leithold, The Calculus with Analytic Geometry, 4th ed., Harper & Row, New York, 1981.

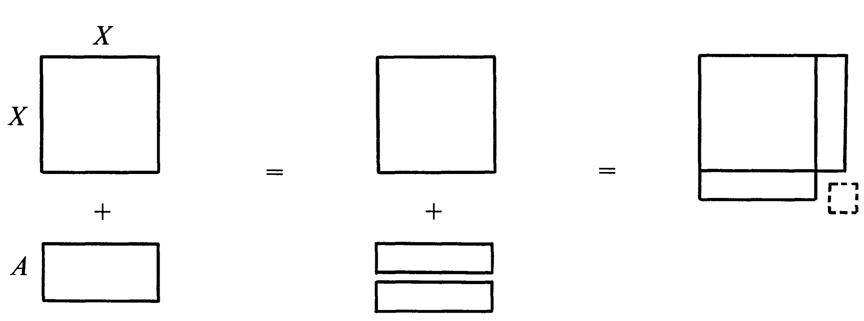
[4] A. I. Markushevich, The Remarkable Sine Functions (translated by Scripta Technica, Inc.), Amer. Elsevier, New York, 1966.

[5] C. L. Siegel, Topics in Complex Function Theory, vol. I: Elliptic Functions and Uniformization Theory (translated by Shenitzer and Solitar), Wiley–Interscience, New York, 1969.

[6] George B. Thomas and Ross L. Finney, Calculus and Analytic Geometry, 5th ed., Addison–Wesley, Reading, 1979.

Proof without words: Completing the square

$$X^2+AX=(X+A/2)^2-(A/2)^2$$



—CHARLES D. GALLANT
St. Francis Xavier University
Antigonish, Nova Scotia

PROBLEMS

Leroy F. Meyers, Editor
G. A. Edgar, Associate Editor
The Ohio State University

Proposals

To be considered for publication, solutions should be mailed before August 1, 1983.

1165. For real numbers a, b with $a<b$ and positive integers n consider the matrix $A_n = (a_{i,j})_{i,j=0,1,\ldots,n}$ where $a_{i,j} = \int_a^b x^{i+j}\,dx$. Prove that

$$\det A_n = \frac{n!\cdot 2^n}{(2n+1)!}(b-a)^{(n+1)^2}\prod_{k=1}^{n}\binom{2k}{k}^{-2}.$$

[*Heinz-Jürgen Seiffert, student, Freie Universität Berlin.*]

1166. Let m and n be nonnegative integers, let c be a positive irrational number, and let $d=1/c$. Prove that

$$\sum_{j=0}^{[m+nc]}[n+1+(m-j)d] = \sum_{j=0}^{[n+md]}[m+1+(n-j)c],$$

where the brackets denote the greatest integer function. [*Clark Kimberling, University of Evansville.*]

1167. Let S be the set of all integers of the form x^2+y^2 where x and y are integers. It is readily seen from Fermat's identity that if $u\in S$ and $v\in S$, then $uv\in S$. Prove that if $m\in S$ and $n\in S$ and $m/n=l$ is an integer, then $l\in S$, **by elementary means*, i.e., without using the well-known characterization of members of S in terms of their factorization into primes. [*Edward T. H. Wang, Wilfrid Laurier University.*]

1168. Let P be a variable point on side BC of triangle ABC. Segment AP meets the incircle of triangle ABC in two points, Q and R, with Q being closer to A. Prove that the ratio AQ/AP is a minimum when P is the point of contact of the excircle opposite A with side BC. [*Stanley Rabinowitz, Merrimack, New Hampshire.*]

We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) will be placed next to a problem number to indicate that the proposer did not supply 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. It is not necessary to submit duplicate copies.

Send all communications to the problems department to Leroy F. Meyers, Mathematics Department, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.

1169. Suppose that two players, P_1 and P_2, are engaged in a best-of-three chess tournament, with P_1 playing the white pieces in the first game. Let p_1 be the probability that P_1 wins when playing the white pieces, and let p_2 be the corresponding probability for P_2. Suppose that $0 < p_1 < 1$ and $0 < p_2 < 1$, and that the probability of a draw is 0.

(a) When is it better for P_1 to purposely lose the first game (rather than play to win) if the loser of the first game plays the white pieces for the remaining game(s)?

(b) Show that it is never to P_1's advantage to purposely lose the first game if the rule is that the loser of any game plays the white pieces in the next game. [*Peter Schumer, student, University of Maryland.*]

Quickies

Solutions to Quickies appear at the conclusion of the Problems section.

Q682. A solid right circular cone stands on its horizontal base. An elliptical plane section, E, of the cone is marked on its surface. V and U are the lowest and highest points of E. Two ants start at V and crawl over the cone to U, one along E and the other along the shortest parth from V to U. Is it possible for their paths to cross before reaching U? [*Robert C. Lyness, Southwold, Suffolk, England.*]

Q683. Find all rational x and y such that $x^2 + 4 = y^2$. [*A. Wilansky, Lehigh University.*]

Q684. Find that factor of $2^{33} - 2^{19} - 2^{17} - 1$ which lies between 1000 and 5000. [*Noam Elkies, student, Stuyvesant High School, New York; member of the* 1982 *U. S. Olympiad Team.*]

Solutions

The Unbiased Coin (continued) **January 1967**

643. Prove that the probability of a match (HH or TT) is $\frac{1}{2}$ if and only if at least one of the coins is unbiased. Generalize to more than two coins. [*Richard L. Eisenman, U. S. Air Force Academy.*]

Solution II: Let n coins fall heads with respective probabilities $p_1, p_2, \ldots, p_n$. Then the probability that exactly m of the n coins fall heads is the coefficient of x^m in the expansion of

$$P(x) = (xp_1 + 1 - p_1)(xp_2 + 1 - p_2) \cdots (xp_n + 1 - p_n).$$

Thus, the probability of getting an even number of heads is given by

$$\frac{P(1) + P(-1)}{2} = \frac{1 + (1 - 2p_1)(1 - 2p_2) \cdots (1 - 2p_n)}{2}.$$

This probability is equal to $\frac{1}{2}$ just when $p_i = \frac{1}{2}$ for some i, i.e., at least one of the coins is fair.

MICHAEL MCGRATH, student
Gunn High School
Palo Alto, California

Solution I, with comments, appeared in vol. 40 (1967), pp. 223–224.

1008. (a) Let $\{a_n\}$ be an increasing sequence of positive integers and let $s_n = a_1 + a_2 + \cdots + a_n$. Show that if $\underline{\lim}\ a_n/n > 2 + \sqrt{2}$, then for all n sufficiently large there exists a perfect square between s_n and s_{n+1}.

(b) Show that if the above conclusion fails and $\underline{\lim}\ a_n/n = 2 + \sqrt{2}$, then $\overline{\lim}\ a_n/n = \infty$. [*Paul Erdös, Hungarian Academy of Science, and Melvyn B. Nathanson, Southern Illinois University at Carbondale.*]

Solution (filled in by the editors): Since $\{a_n\}$ is an increasing sequence of integers, an easy induction shows that

$$a_n \geqslant a_{n-k} + k \qquad \text{for } 0 \leqslant k \leqslant n-1. \tag{1}$$

Adding these n inequalities, we get $na_n \geqslant s_n + n(n-1)/2$, or

$$4na_n - 2n(n-1) - 4s_n \geqslant 0. \tag{2}$$

Let M be the set of all positive integers n such that there is no perfect square between s_{n-1} and s_n, i.e., such that there is no integer between $\sqrt{s_{n-1}}$ and $\sqrt{s_n}$. Thus, if $n \in M$, then $\sqrt{s_n} - 1 \leqslant \sqrt{s_{n-1}} < \sqrt{s_n}$, and so

$$a_n = s_n - s_{n-1} \leqslant s_n - \left(\sqrt{s_n} - 1\right)^2 = 2\sqrt{s_n} - 1.$$

Therefore,

$$4s_n - (a_n + 1)^2 \geqslant 0 \qquad \text{if } n \in M. \tag{3}$$

Combining (2) and (3) and setting $c_n = a_n/n$, we have

$$\begin{aligned} 0 &\leqslant \frac{1}{n^2}\left(4na_n - 2n(n-1) - 4s_n\right) \\ &\leqslant \frac{1}{n^2}\left(4na_n - 2n(n-1) - 4s_n\right) + \frac{1}{n^2}\left(4s_n - (a_n+1)^2\right) \\ &= 4c_n - 2 + \frac{2}{n} - \left(c_n + \frac{1}{n}\right)^2 \\ &= \frac{2}{n}(1 - c_n) - \frac{1}{n^2} - \left(c_n^2 - 4c_n + 2\right) \qquad \text{if } n \in M. \end{aligned} \tag{4}$$

Now from $(c_n - (2 + \sqrt{2}))^2 \geqslant 0$ we obtain

$$c_n^2 - 4c_n + 2 \geqslant 2\sqrt{2}\left(c_n - (2 + \sqrt{2})\right).$$

Combining this with (4) yields

$$0 \leqslant \frac{2}{n}(1 - c_n) - \frac{1}{n^2} - 2\sqrt{2}\left(c_n - (2 + \sqrt{2})\right),$$

or

$$c_n \leqslant \frac{\dfrac{2}{n} - \dfrac{1}{n^2} + 2\sqrt{2}\,(2 + \sqrt{2})}{2\sqrt{2} + \dfrac{2}{n}} \qquad \text{if } n \in M. \tag{5}$$

(a) Assume $\underline{\lim}\ c_n > 2 + \sqrt{2}$. The right side of (5) converges to $2 + \sqrt{2}$ as $n \to \infty$. Therefore (5) holds for only finitely many n, and thus M is a finite set, which is the required conclusion.

(b) Now assume $\underline{\lim}\, c_n = 2 + \sqrt{2}$ and M is infinite. We will show that $\overline{\lim}\, c_n = +\infty$. To do this, it suffices to show that $\overline{\lim}(s_n/n^2) = +\infty$. In fact, we will show that $\overline{\lim}(s_n/n^2) > d$ for each fixed $d > 1$.

By (5) we have $\overline{\lim}_{n \in M} c_n \leqslant 2 + \sqrt{2}$, and so, in fact, $\lim_{n \in M} c_n = 2 + \sqrt{2}$. Hence

$$\lim_{n \in M}\left(\frac{2}{n}(1 - c_n) - \frac{1}{n^2} - \left(c_n^2 - 4c_n + 2\right)\right) = 0,$$

and so the squeeze theorem applied to (4) yields

$$\lim_{n \in M}\frac{1}{n^2}(4na_n - 2n(n-1) - 4s_n) = 0,$$

or

$$\lim_{n \in M}\frac{1}{n^2}\sum_{k=0}^{n-1}(a_n - a_{n-k} - k) = 0. \tag{6}$$

For $n \in M$ and $n \geqslant d$, let m be the greatest integer in n/d. Since, by (1), each summand in (6) is nonnegative, the squeeze theorem yields

$$\lim_{n \in M}\frac{1}{n^2}\sum_{k=n-m}^{n-1}(a_n - a_{n-k} - k) = 0,$$

or

$$\lim_{n \in M}\frac{1}{n^2}\left(ma_n - s_m - \frac{n(n-1)}{2} + \frac{(n-m)(n-m-1)}{2}\right) = 0,$$

or

$$\lim_{n \in M}\left(\frac{m}{n}c_n - \frac{1}{n^2}s_m - \frac{1}{2} + \frac{1}{2n} + \frac{\left(1 - \frac{m}{n}\right)\left(1 - \frac{m}{n} - \frac{1}{n}\right)}{2}\right) = 0.$$

Since $\lim_{n \in M}(m/n) = 1/d$, we have

$$\lim_{n \in M}\left(\frac{1}{d}c_n - \frac{1}{d^2}\frac{s_m}{m^2} - \frac{1}{2} + 0 + \frac{\left(1 - \frac{1}{d}\right)\left(1 - \frac{1}{d} - 0\right)}{2}\right) = 0,$$

and multiplication by d^2 yields

$$\lim_{n \in M}\left(dc_n - \frac{s_m}{m^2} - d + \frac{1}{2}\right) = 0.$$

Hence

$$\lim_{n \in M}\frac{s_m}{m^2} = d(2 + \sqrt{2}) - d + \frac{1}{2} > d,$$

and so $\overline{\lim}(s_n/n^2) > d$, as was to be shown.

PAUL ERDŐS
Hungarian Academy of Science

MELVYN B. NATHANSON
Southern Illinois University at Carbondale

A Matrix Shifts Left — January 1982

1138. Let X be a nonsingular matrix with columns $X_1, X_2, \ldots, X_n$. Let $Y=[X_2, X_3, \ldots, X_n, \mathbf{0}]$. Show that the matrices $A = YX^{-1}$ and $B = X^{-1}Y$ have rank $n-1$ and have only 0 for eigenvalues. Conversely, show that every $n \times n$ matrix A of rank $n-1$ and with only 0 for eigenvalues can be written $A = YV$ for some nonsingular $X = V^{-1}$ and Y defined as above. [*John Z. Hearon, National Institutes of Health.*]

Solution: Let $J=[a_{ij}]$ be the $n \times n$ Jordan matrix where $a_{ij}=1$ if $i=j+1$ and $a_{ij}=0$ otherwise. The rank of J is $n-1$ and its only eigenvalues are 0.

If X is the nonsingular matrix with columns $X_1, \ldots, X_n$, and if $Y=[X_2, X_3, \ldots, X_n, \mathbf{0}]$, then $Y=XJ$. If $A = YX^{-1}$, then $X^{-1}AX=J$, so that J is the Jordan form for A. Moreover, if $B = X^{-1}Y$, then $B=J$. It follows that both A and B have rank $n-1$ with only 0 for eigenvalues.

Conversely, if A has rank $n-1$ with only 0 for eigenvalues, then its Jordan form is J. Thus there exists a nonsingular matrix X such that $X^{-1}AX=J$. Then $A = YX^{-1}$ where $Y=XJ$, and the columns of X and Y are as above.

HARRY SEDINGER
St. Bonaventure University

Also solved by Chico Problem Group, William H. Gustafson, Konrad J. Heuvers, W. C. Pye & Garry L. Walls, Donald W. Robinson, J. M. Stark, Yan-Loi Wong (Hong Kong), and the proposer.

Irreducible Polynomial — January 1982

1139. Let $p(x)$ be a polynomial with rational coefficients and suppose $p(x)$ is irreducible over the rationals. Let α be a complex number such that $p(\alpha)=0$. Then we know $p(x)=(x-\alpha)q(x)$, where $q(x)$ is a polynomial with complex coefficients. The leading coefficient of $q(x)$ is rational. Can any of the other coefficients of $q(x)$ be rational numbers? [*Roger L. Creech, East Carolina University.*]

Solution: Put $p(x)=p_0x^n+p_1x^{n-1}+\cdots+p_n$ (where $p_0 \neq 0$) and $q(x)=q_0x^{n-1}+\cdots+q_{n-1}$. From $p(x)=(x-\alpha)q(x)$ we obtain, by comparing coefficients and re-arranging terms (or by using synthetic division),

$$q_r=p_0\alpha^r+p_1\alpha^{r-1}+\cdots+p_r \qquad (r=0,1,\ldots,n-1).$$

Since α is a zero of a polynomial of degree n which is irreducible over $\mathbf{Q}$, the numbers $1, \alpha, \alpha, \ldots, \alpha^{n-1}$ are linearly independent over $\mathbf{Q}$. So, for $r \geqslant 1$, the coefficient q_r cannot be rational unless $p_0=p_1=\cdots=p_{r-1}=0$. Since $p_0 \neq 0$, only q_0 is rational.

P. B. KRONHEIMER
Merton College, Oxford, England

Also solved by Duane Broline, Stephen D. Bronn, Alan Edelman (student), Chico Problem Group, G. A. Heuer, Steve Kahn, L. Kuipers (Switzerland), Henry S. Lieberman, Daniel A. Rawsthorne, J. M. Stark, and the proposer. There was one incorrect solution.

A Logarithmic and Exponential Inequality — March 1982

1140. Show that, if $0<x<y$, then

$$x+\ln(e^y-y-1) \geqslant x^{3/4}(x+y)^{1/4}+\ln y\left(\sqrt{x(x+y)}-x\right).$$

[*Mihály Bencze, Braşov, Romania.*]

Solution: We shall prove the slightly stronger inequality: If x and y are positive, then

$$x+\ln(e^y-y-1)\geqslant x^{2/3}(x+y)^{1/3}+\ln y\left(\sqrt{x(x+y)}-x\right).$$

Since $\binom{k+2}{2}\leqslant 3^k$ for $k=0,1,2,\ldots$, and $y>0$, we have

$$\frac{2(e^y-y-1)}{y^2}=\sum_{k=0}^{\infty}\frac{y^k}{k!\binom{k+2}{2}}>e^{y/3}>e^{y/4}.$$

Thus, since $y/x>0$ and $0<\frac{1}{3}<1$, we have, by Bernoulli's inequality,

$$\begin{aligned}x+\ln(e^y-y-1)&>x\left(1+\frac{y}{3x}\right)+\ln\frac{y^2}{2}\\&>x\left(1+\frac{y}{x}\right)^{1/3}+\ln\frac{y^2}{1+\sqrt{1+\frac{y}{x}}}\\&=x^{2/3}(x+y)^{1/3}+\ln y\left(\sqrt{x(x+y)}-x\right).\end{aligned}$$

J. S. FRAME
Michigan State University

Also solved by Nick Franceschine III (generalized), V. D. Mascioni (Switzerland), Otto G. Ruehr, J. M. Stark, and the proposer. There were two incorrect solutions.

Sums of Independent Binomial Random Variables **March 1982**

1141. Let $X_1, X_2, \ldots, X_k$ be independent binomial random variables with parameters (n_1, p_1), $(n_2, p_2), \ldots, (n_k, p_k)$, respectively. Let $S=X_1+X_2+\cdots+X_k$, $n=n_1+n_2+\cdots+n_k$, and $p=\max(p_1, p_2, \ldots, p_k)$. If j is a fixed integer chosen from $0,1,2,\ldots,n$, show that $P[S\geqslant j]\leqslant P[X\geqslant j]$, where X is a binomial random variable with parameters (n,p). (Recall that a binomial random variable with parameters (n,p) is a random variable X taking only the values $0,1,\ldots,n$ and satisfying $P[X=k]=\binom{n}{k}p^k(1-p)^{n-k}$.) [*Joe Dan Austin, Rice University.*]

Solution: It suffices to prove the following more general result:

THEOREM. *Let $X_1,\ldots,X_n$ and $X_1^*,\ldots,X_n^*$ be two sequences of independent* 0-1 *Bernoulli trials with $P\{X_i=1\}=p_i$, $P\{X_i^*=1\}=p_i^*$, $p_i\leqslant p_i^*$, and $S_n=X_1+\cdots+X_n$, $S_n^*=X_1^*+\cdots+X_n^*$. Then*

$$P\{S_n\geqslant j\}\leqslant P\{S_n^*\geqslant j\}$$

for all $0\leqslant j\leqslant n$.

Proof. (a) By induction. By hypothesis the theorem holds for $n=1$, so assume it holds for $n-1$. Then

$$\begin{aligned}P\{S_n\geqslant j\}&=P\{S_{n-1}\geqslant j\}+p_nP\{S_{n-1}=j-1\}\\&\leqslant P\{S_{n-1}\geqslant j\}+p_n^*P\{S_{n-1}=j-1\}\\&=P\{S_{n-1}\geqslant j\}+p_n^*[P\{S_{n-1}\geqslant j-1\}-P\{S_{n-1}\geqslant j\}]\\&=(1-p_n^*)P\{S_{n-1}\geqslant j\}+p_n^*P\{S_{n-1}\geqslant j-1\}\\&\leqslant(1-p_n^*)P\{S_{n-1}^*\geqslant j\}+p_n^*P\{S_{n-1}^*\geqslant j-1\}\\&=P\{S_n^*\geqslant j\}.\end{aligned}$$

(b) Alternatively, simply note that the first two lines in proof (a) prove the result when $p_1=p_1^*,\ldots,p_{n-1}=p_{n-1}^*$, $p_n\leqslant p_n^*$ (i.e., when S_n and S_n^* differ in only one summand), and then "switch" X_i^* for X_i one at a time.

(c) The following proof has the advantage of providing a simple illustration of the use of "coupling" methods in probability.

Let $\Omega = \{\omega = (\omega_1, \dots, \omega_n) \in [0,1]^n\}$ be endowed with the usual uniform measure, let I_A denote the indicator function of a set A (i.e., $I_A(\omega) = 1$, $\omega \in A$ and $I_A(\omega) = 0$, $\omega \notin A$), and let

$$X_i(\omega) = I_{[0, p_i]}(\omega_i), \; X_i^*(\omega) = I_{[0, p_i^*]}(\omega_i).$$

Then clearly $\{X_i\}$, $\{X_i^*\}$ satisfy the hypotheses of the theorem, and $X_i(\omega) = 1 \Rightarrow X_i^*(\omega) = 1$ for *all* ω; hence $S_n(\omega) \leqslant S_n^*(\omega)$ for all ω, and $P\{S_n \geqslant j\} \leqslant P\{S_n^* \geqslant j\}$.

(d) Given two random variables X and Y, say that X is *stochastically bounded* by Y $(X < Y)$ if $P\{X \geqslant t\} \leqslant P\{Y \geqslant t\}$ $(\Leftrightarrow F_Y(t) \leqslant F_X(t)$, where $F_X(t) = P\{X \leqslant t\}$ is the cumulative distribution function of X). Then the theorem is a special case of the following, more general, result:

stochastic boundedness is preserved under sums of independent random variables,

i.e., if X, Y and X^*, Y^* are independent pairs of random variables such that $X < X^*$, $Y < Y^*$, then $X + Y < X^* + Y^*$.

Proof. (i)
$$\begin{aligned} F_{X+Y}(z) &= \int F_X(z-y)\, dF_Y(y) \\ &\geqslant \int F_{X^*}(z-y)\, dF_Y(y) \\ &= F_{X^*+Y}(z) \geqslant F_{X^*+Y^*}(z), \end{aligned}$$

where the second inequality follows for the same reasons as the first, or by switching Y^* for Y.

(ii) Let $V\colon \Omega \to [0,1]$ be a uniformly distributed random variable, define $F_X^{-1}(y) = \inf\{x : y \leqslant F(x)\}$, and observe that $F_X^{-1}(V) \sim X$. Thus if V_1, V_2 are independent uniforms on $[0,1]$, then

$$\begin{aligned} &F_X^{-1}(V_1) \leqslant F_{X^*}^{-1}(V_1), \; F_Y^{-1}(V_2) \leqslant F_{Y^*}^{-1}(V_2) \\ &\quad \Rightarrow F_X^{-1}(V_1) + F_Y^{-1}(V_2) \leqslant F_{X^*}^{-1}(V_1) + F_{Y^*}^{-1}(V_2) \\ &\quad \Rightarrow X + Y < X^* + Y^*. \end{aligned}$$

For information about stochastic boundedness and its relationship to the monotone likelihood ratio property, see E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, 1959, pp. 68–75, 111, 329–333, 343; and D. Gilat, *Monotonicity of a power function: an elementary probabilistic proof*, Amer. Statist., vol. 31 (1977), pp. 91–93.

SANDY L. ZABELL
Northwestern University

Also solved by Victor Hernandez (Spain), Kumar Joag-Dev, Bruce R. Johnson (Canada), K. Aswath Rao, J. M. Stark, and the proposer.

Both Johnson and the proposer remarked that if W is a binomial random variable with parameters $(n, \min(p_1, \dots, p_n))$, then $P[W \geqslant j] \leqslant P[S \geqslant j]$. The proposer showed that $P[X \geqslant j] = P[S \geqslant j]$ if and only if either $j = 0$ or all p_i are equal.

Unit Resistors and Fibonacci Numbers **March 1982**

1142. If a two-terminal series-parallel circuit built with n unit resistors has total resistance p/q, where $(p, q) = 1$, prove that p does not exceed the $(n+1)$th Fibonacci number. [*Jeremy D. Primer, student, Princeton University.*]

Editor's comment. *Eric S. Rosenthal* observed that this problem was solved as part of the solution of problem **E2459** in *The American Mathematical Monthly*, vol. 81 (1974), p. 170, and vol. 82 (1975), pp. 178–182. The solution presented there is essentially the same as those submitted here, using induction with the strengthened hypothesis $p \leqslant F_{n+1}$, $q \leqslant F_{n+1}$, and $p+q \leqslant F_{n+2}$, as well as the identity $F_a F_b + F_{a+1}F_{b+1} = F_{a+b+1}$.

A more general problem, but without complete solution, can be found as problem **3430** in the *Monthly*, vol. 37 (1930), p. 261, and vol. 38 (1931), pp. 172–173. The problem of finding a circuit with resistance 335/133 (an excellent approximation to π) is the *Monthly* problem **E2429** cited above, and has also appeared in *BYTE* (February 1980, p. 16; August 1980, p. 20; January 1981, p. 16; and May 1981, p. 268).

Solved by Dan Rawsthorne, and by Michael Larsen & the proposer. Partial solutions by Anders Bager (Denmark), Mark F. Kruelle, and James Morrow.

Boolean Rings—Again! **March 1982**

1143. (a) Show that a ring R is a Boolean ring if and only if $x^{20} = x$ for all elements x of R.

(b*) For which positive integers k is it true that a ring R satisfying $x^k = x$ for all x in R must be Boolean? [*Thomas Hand, Indiana State University, Terre Haute.*]

Editor's note. Many correspondents remarked that the problem is well known. In fact, it has appeared as problem **5972** in *The American Mathematical Monthly* (proposed, vol. 81 (1974), p. 534; solved, vol. 83 (1976), pp. 66–67), as well as problem **E2536** (proposed, vol. 82 (1975), p. 521; solved, vol. 83 (1976), pp. 657–658). Related problems are **1019** (this MAGAZINE, proposed, vol. 50 (1977), p. 164; solved, vol. 52 (1979), p. 50) and **6284** (*Monthly*, proposed, vol. 86 (1979), p. 869; solved, vol. 89 (1982), pp. 135–136). Many properties of rings for which there is an integer $k > 1$ such that $x^k = x$ for all x in the ring (called J-rings, after Jacobson) can be found in the papers: R. & C. Ayoub, "On the commutativity of rings," *American Mathematical Monthly*, vol. 71 (1964), pp. 267–271; and Jiang Luh, "On the structure of J-rings," *ibid.*, vol. 74 (1967), pp. 164–166.

All but one of the complete proofs submitted in answer to part (b) were similar to the proofs published for the *Monthly* problems cited above. The proofs submitted for part (a) only, or for a part of part (b), were more elementary.

Solution: (a) The "only if" part is trivial, so suppose $x^{20} = x$ for all x in the ring R. For every $x \in R$ we have $-x = (-x)^{20} = x^{20} = x$, so $2x = 0$. Of the binomial coefficients $\binom{20}{r}$ for $1 \leqslant r \leqslant 19$, only $\binom{20}{4}$ and $\binom{20}{16}$ are odd, and so for every $x \in R$ we have

$$x^7 + x^{15} = (x^7 + x^{15})^{20} = x^{140} + x^{28}x^{240} + x^{112}x^{60} + x^{300} = x^7 + x^2 + x + x^{15},$$

giving $x^2 = x$.

A. J. DOUGLAS & G. T. VICKERS
The University of Sheffield (England)

(b) The identity

$$x^k = x \text{ for all } x \text{ in } R \qquad (*)$$

implies that R is Boolean if and only if k is even and for no $n > 1$ does $2^n - 1$ divide $k - 1$.

First, if k is odd, then the non-Boolean ring Z_3 satisfies $(*)$, and if $2^n - 1$ divides $k - 1$ for $n > 1$, then the non-Boolean ring $GF(2^n)$, the Galois field of 2^n elements, satisfies $(*)$.

We now suppose that k is even, that $2^n - 1 \nmid k - 1$, and that R satisfies $(*)$. As in part (a), for every $x \in R$ we have $-x = (-x)^k = x^k = x$, so that $2x = 0$. Hence R has characteristic 2.

Consider a fixed $y \in R$ and let S be the subring of R generated by y. Since $y^{k-1}y = y$, y^{k-1} is an identity element for S. Then S is a finite-dimensional vector space over Z_2 spanned by 1 $(= y^{k-1})$, $y, y^2, \ldots, y^{k-2}$. For each w in S let T_w be the linear transformation on S defined by

$T_w(z) = wz$ for $z \in S$. In particular, $T_y(z) = yz$ and for every polynomial p over Z_2 we have $T_{p(y)} = p(T_y)$. Hence the polynomial $t^k - t = t^k + t \in Z_2[t]$ annihilates T_y. Since k is even,

$$\frac{d}{dt}(t^k + t) = kt^{k-1} + 1 = 1$$

is relatively prime to $t^k + t$, so that $t^k + t$ has distinct roots in K, the algebraic closure of Z_2. Hence T_y is diagonalizable over K, with matrix $[T_y] = \text{diag}[a_1, \ldots, a_n]$ with respect to a basis of S consisting of eigenvectors of T_y, where the a_i are in K. Then for every polynomial p we have $[T_{p(y)}] = p([T_y]) = \text{diag}[p(a_1), \ldots, p(a_n)]$. Let F be the subring of K generated by a_1. Since $(p(y))^k = p(y)$ for all polynomials p, F satisfies $(*)$. But K is a field, so that if $z \in F$ and $z \neq 0$, then $1/z = z^{k-2} \in F$. Thus F is a subfield of K. Since F satisfies $(*)$, F is finite and so $F = GF(2^n)$ for some n. Since the nonzero elements of F form a cyclic group, $2^n - 1$ must divide $k - 1$, which is contrary to hypothesis unless $n = 1$. Hence $F = GF(2) = Z_2$, so that $a_1^2 = a_1$; similarly, $a_i^2 = a_i$ for all i. Then $T_{y^2} = (T_y)^2 = T_y$, so that $y^2 = T_{y^2}(1) = T_y(1) = y$. Since y was any element of R, R is Boolean.

THOMAS JAGER
Calvin College

Complete solutions or references to part (b) *were submitted by Anders Bager (Denmark), Paul J. Campbell & Stephen J. Curran, Lee Erlebach, H. P. A. Künzi (Switzerland), John J. Martinez, Paul S. Peck, Daniel M. Rosenblum, Alasdair Urquhart (Canada), William P. Wardlaw, Gregory P. Wene, and A. Wilansky.*

Solutions to part (a) *or partial solutions to part* (b) *were submitted by L. Richard Duffy (student), Marshall Fraser, Rodica Simion & Frank W. Schmidt, and the proposer.* There was one seriously incomplete solution.

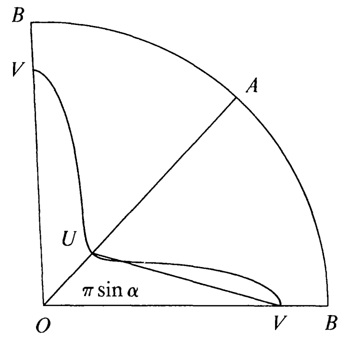
Answers

Solutions to the Quickies which appear near the beginning of the Problems section.

Q682. Yes. Let O be the vertex and α the semivertical angle of the cone. On the plane development of the cone (see the figure), if $\angle VUO > 90°$ [i.e., $OU < OV\cos(\pi \sin\alpha)$], the straight line VU must cut the first ant's path, for the latter meets OA and OB at right angles. [The condition can also be expressed as

$$\tan^2\left(\frac{\pi}{2}\sin\alpha\right) < (\tan\alpha)(\tan\beta) < 1,$$

where β is the angle between the plane and the base—but this is not a Quickie! This problem appeared in the 1982 British Mathematical Olympiad; see *Crux Mathematicorum*, vol. 8 (1982) 135. *Ed.*]



Q683. Let $z = (y + x)/2$. Then $1/z = (y - x)/2$, so that $x = z - 1/z$. Conversely, if $x = z - 1/z$, then $x^2 + 4 = (z + 1/z)^2$.

Q684. $2^{33} - 2^{19} - 2^{17} - 1 = (2^{11})^3 - (2^6)^3 - 1^3 - 3 \cdot 2^{11} \cdot 2^6 \cdot 1$, which is of the form $x^3 - y^3 - z^3 - 3xyz$, and so has $x - y - z$ as a factor. Thus the requisite factor is $2^{11} - 2^6 - 1 = 1983$. (The complete factorization is $3^3 \cdot 13 \cdot 661 \cdot 37021$.)

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.

Marty, Martin E., *The alluring postulate*, The Christian Century, 99:39 (8 December 1982) 1271.

Have you ever wondered how to make love to a postulate? This column is just what you need to enliven a dull faculty or business meeting, or to distribute to a bored class yearning for spring break.

Pomerance, Carl, *The search for prime numbers*, Scientific American 247:6 (December 1982) 136-147, 178.

Splendid exposition of the history of primality-testing, up through the 1980 breakthrough by Adleman and Rumely and its improvements. (On p. 136, third column, the result attributed to Chin Jing-run should say "all large enough even numbers can be expressed as the sum of a prime number and a number that is either prime or the product [not "sum"] of two primes.")

Steen, Lynn Arthur, *Twisting and turning in space; classifying shapes and surfaces in new dimensions suggests the possibility of at last achieving a unified geometry*, Science News 122 (17 July 1982) 42-44.

Recent progress on manifolds appears to signal the nearing of the end of the effort to classify them. The last unproved case of the Poincaré conjecture is dimension three: that every three-dimensional object whose homotopy agrees with the sphere's is topologically equivalent to a sphere.

Morris, Scott, *Games: Fours, the sons of Rubik, and a ΨΦ film quiz*, Omni 5:1 (October 1982) 200-201.

Photos and description of recent products from the cube puzzle world, including the "world's hardest cube puzzle (so far)." (Since your local store probably doesn't have all of these, Morris provides a mail-order address.)

Vere-Jones, David, *Dialogue with a Soviet mathematics teacher, Parts 1 and 2*, New Zealand Mathematics Magazine, 18 (1982) 77-85, 19 (1982) 4-8.

Discussion with a recent emigrant who taught school mathematics in the Soviet Union for over 20 years. Although the Soviet system has virtually all students cover in 10 years the mathematics through calculus, there are problems with the high degree of "theory" and how well it is absorbed; so that "even a qualitative understanding of the material in the course for the 9th and 10th classes is beyond the reach of the weaker or even average pupils."

Carlos Ramirez and Albert H. Woods, Inc., "Calculating to Computing...the Dawn of the Information Age," new exhibition presented by IBM at the Museum of Science and Industry, Chicago.

A major new permanent exhibit on the history of computing, opened at the end of October in Chicago. Originally conceived as IBM's own technical record of its innovations in computer technology, the exhibit has three aspects: first, a long and varied wall tableau of the history of computing, including reproductions of famous documents and apparatus (e.g., Pascal's calculator); second, half a dozen glass-enclosed computer rooms from different IBM eras; and third, ten IBM pre-programmed microcomputers. The technical historical consultant was Bernard Cohen (Harvard). What was removed to make room for this exhibit? You guessed it--the IBM Mathematica exhibit.

Peterson, Ivars, *Can you count on your computer?*, Science News 122 (31 July 1982) 72-75.

The quality of arithmetic performed by calculators and computers still leaves much to be desired; for example, early IBM Personal Computers displayed 0.001 as the result of 0.1 divided by 10, and a leading manufacturer's natural log function is off by 10% for arguments near 1. But there is hope for progress in the proposed standard for binary floating-point arithmetic, which is implemented in the Intel 8087 chip. The article has a humorous horror story of the true experience of an aeronautical engineer whose simulated plane kept crashing--on computer after computer--solely because of erroneous computer arithmetic.

McClain, Ellen Jaffe, *Do women resist computers?*, Popular Computing 2:3 (January 1983) 66-78.

Upbeat popular article that covers all the angles and concludes "quite the opposite"--in fact, that women's "skills and enthusiasm may change the course of computing in the 80's." One pattern does emarge: men have more fun with computers, while women tend to view the computer as a tool, a means to an end.

Weizenbaum, Joseph, *Playing with violence*, Fellowship 48:9 (September 1982) 3-4, 15.

The author of *Computer Power and Human Reason* here cites the danger of very large computer systems (such as used in the North American Defense Command): they aren't understood by anyone, so that when "an error" occurs, it isn't possible to find *the* error. He also decries the kind of participation computer games foster: "Children are able to shoot the torpedoes themselves and repeatedly experience the thrill of having killed a great many people." Finally, he questions what lessons students are expected to learn, versus what they actually learn, from computers in the classroom.

Kolata, Gina, *Does Gödel's theorem matter to mathematics? The recent discovery of two natural but undecidable statements indicates that Gödel's theorem is more than just a logician's trick*, Science 218 (19 November 1982) 779-780.

Harvey Friedman (Ohio State) has found yet another statement undecidable in Peano arithmetic, and this one is even more "natural" than earlier ones found by Paris and Harrington a few years ago. The discovery increases the willingness of many mathematicians to believe that witnesses for Gödel's theorem on undecidability may include important results, such as Fermat's Last Theorem.

Hellman, Geoffrey, *How to Gödel a Frege-Russell: Gödel's incompleteness theorem and logicism*, Nous 15 (1981) 451-468.

Formalism runs afoul of Gödel's first incompleteness theorem; how fares logicism? The author argues that so long as logic is taken to be formalizable, logicism is incompatible with Gödel's *second* incompleteness theorem.

Dobbs, David, and Hanks, Robert, A Modern Course on the Theory of Equations, Polygonal Publ. House, 1980; viii + 216 pp.

By 1960 courses in theory of equations had been supplanted in the U.S. by courses in abstract algebra. This book recalls the old days from today's perspective by offering a thoroughly algebraic approach to the theory of polynomials, accompanied by computer-era considerations about numerically approximating roots. The result is a delightful mathematically-oriented (as opposed to a "discrete structures" computing-oriented) alternative to the "traditional" abstract algebra of groups-rings-fields.

Hilton, Peter, and Pedersen, Jean, Fear No More: An Adult Approach to Mathematics, Addison-Wesley, 1983; viii + 281 pp.

Splendid rethinking of the purposes, content, and presentation of arithmetic for mature adults, in a text suitable for self-study. "...[A]rithmetical training is, traditionally, not accompanied by an education in mathematics.... [M]athematics involves, in an essential way, abstraction, analytical reasoning, computation, and interpretation." Calculators render most rote computation training obsolete, which is good, because computation has always been the least human and least mathematical aspect of mathematics. The "back-to-basics" movement is fundamentally anti-mathematics and anti-education; understanding *must* accompany skill. The authors treat topics in relation to their importance for today's non-scientist adult: the arithmetic of estimation and approximation receives a prominent place. Two subsequent volumes will treat geometry-algebra-trigonometry and the calculus.

Gardiner, A., Infinite Processes: Background to Analysis, Springer-Verlag, 1982; viii + 306 pp, $28.

Provides the logical and historical motivation a student needs for an appreciation of advanced calculus. This book critically examines the implicit infinite processes usually glossed over in elementary mathematics through calculus: infinite decimals, length, area, volume, functions. Exercises continue and reinforce the thinking the book encourages. Used either as a prologue to, or as a supplement in, advanced calculus, this book would be extremely valuable.

Klambauer, Gabriel, Problems and Propositions in Analysis, Dekker, 1979; vii + 456 pp, $32 (P).

Over 500 problems with solutions are assembled here, divided roughly equally among the headings arithmetic/combinatorics, inequalities, sequences/series, and real functions. The problems tend to be a little more elementary than Putnam or *Monthly* problems. Each solution appears immediately after the problem statement, a feature which tends to discourage work in favor of reading.

Gelbaum, Bernard, Problems in Analysis, Springer-Verlag, 1982; vii + 228 pp, $28.

Over 500 problems, based on real analysis, measure theory, elementary topology and some theory of topological vector spaces. A bibliography, glossary, and symbol glossary are included.

Trivedi, K. S., Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Prentice-Hall, 1982; x + 624 pp.

One-year introduction to probability, stochastic processes, and statistics (in that order of emphasis and sequence of coverage). Specifically designed for an audience of senior and graduate students of computer science, electrical/computer engineering, reliability engineering, and applied mathematics. All examples and exercises are drawn from the realm of computer hardware, software, and systems design and performance. Traditional statistics topics, as well as proofs in general, are de-emphasized. Solutions manual available.

NEWS & LETTERS

MORE, OLD, GOATS

Since the publication of "A Tale of Two Goats" (this *Magazine*, September 1982), some readers have written to share additional references with me. Professor John Tierney used the goat problem in his Calculus (4th Edition), Allyn and Bacon, p. 355. Professor Harold Dorwart directed me to "Involutes of a circle and a pasturage problem," *Amer. Math. Monthly*, 28 (1921) 328-329. The author "Arc." of this article found the goat problem (outside the fence) proposed in *The Ladies Diary* for 1748! So far, this is the oldest example of mathematical tethering I know of.

Marshall Fraser
San Francisco, CA 94131

Editor's note: In a surprising coincidence, The Canadian Surveyor *posed the problem anew for its readers at the same time Fraser's note appeared in this* Magazine. *Old goats never die.*

MICROCOMPUTER CONFERENCE

The 6th annual SPRING microCOMPUTER SHOW & TELL CONFERENCE will be held at the University of Oklahoma in Norman, OK, on Saturday, May 21, 1983. The purpose of the conference is to enable microcomputer buffs to share hardware, software, and state-of-the-art ideas.

There will be two 45-minute talks and thirty 5-minute talks, interspersed with 30-minute SHOW & TELL periods during which speakers will be at their computers to demonstrate, discuss, and answer questions about their 5-minute presentations.

Prizes will be awarded after each group of six 5-minute presentations. An on-the-spot programming contest, with prizes, will be held. Last year's contestants had 30 minutes to analyze, program, debug, and run the following problem. *Determine all positive integers (whole numbers) N such that $S = N * N$ contains each of the nine digits 1,2,3,4,5,6,7,8,9 exactly once.*

Provision has been made for computer buffs not in actual attendance to also participate in the 1983 conference. Original microcomputer programs may be submitted for publication consideration in the *Conference Proceedings* and for a prize competition.

For additional information, application forms, and directions for submitting programs, send a self-addressed, stamped envelope to:

Dr. Richard V. Andree
601 Elm, Room 423
Norman, OK 73019

SHORT COURSE ON DATA STRUCTURES

The data structures commonly used in computer science (e.g., linked-lists, trees, graphs, etc.) and their applications will be the focus of a short course sponsored by the Ohio Section of the MAA and Denison University with support from the GTE Corp., to be held June 13-July 1, 1983.

The primary objective of the course, taught by Dr. Zaven A. Karian, is to enable mathematics faculty, who already know how to program, to teach an undergraduate course on data structures similar to CS7 of the ACM Curriculum 78 recommendations. The Pascal programming language will be used as the vehicle for algorithm implementation; those who are not fluent in Pascal should have sufficient expertise in another general-purpose programming language (e.g., BASIC or FORTRAN). Costs are $450 for room and board. A grant from the GTE Corp. will cover all expenses related to instruction and use of facilities.

Applications are due March 25, 1983; for further information, contact:

Dr. Zaven A. Karian
Mathematical Sciences Dept.
Denison University
Granville, Ohio 43023
(614) 587-6563

WORKSHOPS IN APPLICABLE MATH

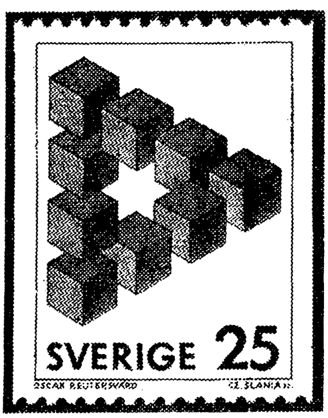
The MD-DC-VA Section of the MAA will sponsor two five-day workshops this June on the Eastern Shore of Maryland. The first, *Microcomputer Graphics*, 13-17 June, will be led by Dr. G. J. Porter of the University of Pennsylvania. The second, *Linear Algebra and the Microcomputer*, 20-24 June, will be led by Dr. Gareth Williams of Stetson Univ., Florida.

The 1983 workshops will be the 8th year in this successful series. They are specifically intended for college teachers. The total cost, including room and board, is $185 for each workshop. For more infomation, contact:

Dr. B. A. Fusaro
Dept. Math. Sciences
Salisbury State College
Salisbury, Maryland 21801
(301) 543-6465

FOR STAMP BUFFS

Sweden has issued a set of three stamps depicting "impossible figures" designed by Oscar Reutersvard.



WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

The 43rd Annual Putnam competition was held December 4, 1982. There were 2,024 student competitors at colleges and universities in the U.S. and Canada who were challenged by the twelve problems that are published here for your information and pleasure. Hints and solutions will appear in our next issue.

A-1. Let V be the region in the cartesian plane consisting of all points (x,y) satisfying the simultaneous conditions

$$|x| \le y \le |x| + 3 \text{ and } y \le 4.$$

Find the centroid $(\bar{x},\bar{y})$ of V.

A-2. For positive real x, let

$$B_n(x) = 1^x + 2^x + 3^x + \ldots + n^x.$$

Prove or disprove the convergence of

$$\sum_{n=2}^{\infty} \frac{B_n(\log_n 2)}{(n \log_2 n)^2}.$$

A-3. Evaluate

$$\int_0^{\infty} \frac{\text{Arctan } (\pi x) - \text{Arctan } x}{x}\, dx .$$

A-4. Assume that the system of simultaneous differential equations

$$y' = -z^3, \quad z' = y^3$$

with the initial conditions $y(0) = 1$, $z(0) = 0$ has a unique solution $y = f(x)$, $z = g(x)$ defined for all real x. Prove that there exists a positive constant L such that for all real x,

$$f(x + L) = f(x), \quad g(x + L) = g(x).$$

A-5. Let a, b, c, and d be positive integers and

$$r = 1 - \frac{a}{b} - \frac{c}{d} \cdot$$

Given that $a + c \leq 1982$ and $r > 0$, prove that

$$r > \frac{1}{1983^3} \cdot$$

A-6. Let σ be a bijection of the positive integers, that is, a one-to-one function from $\{1, 2, 3, \ldots\}$ onto itself. Let $x_1, x_2, x_3, \ldots$ be a sequence of real numbers with the following three properties:

(i) $|x_n|$ is a strictly decreasing function of n;

(ii) $|\sigma(n) - n| \cdot |x_n| \to 0$ as $n \to \infty$;

(iii) $\lim_{n\to\infty} \sum_{k=1}^{n} x_k = 1$.

Prove or disprove that these conditions imply that

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_{\sigma(k)} = 1 \,.$$

B-1. Let M be the midpoint of side BC of a general ΔABC. Using the *smallest possible* n, describe a method for cutting ΔAMB into n triangles which can be reassembled to form a triangle congruent to ΔAMC.

B-2. Let $A(x,y)$ denote the number of points (m,n) in the plane with integer coordinates m and n satisfying

$$m^2 + n^2 \leq x^2 + y^2 \quad .$$

Let

$$g = \sum_{k=0}^{\infty} e^{-k^2} \,.$$

Express

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x,y)\, e^{-x^2-y^2}\, dx\, dy$$

as a polynomial in g.

B-3. Let p_n be the probability that $c + d$ is a perfect square when the integers c and d are selected independently at random from the set $\{1,2,3,\ldots,n\}$. Show that

$$\lim_{n\to\infty} (p_n \sqrt{n})$$

exists and express this limit in the form $r(\sqrt{s} - t)$ where s and t are integers and r is a rational number.

B-4. Let $n_1, n_2, \ldots, n_s$ be distinct integers such that

$$(n_1 + k)(n_2 + k)\cdots(n_s + k)$$

is an integral multiple of

$$n_1 n_2 \cdots n_s$$

for every integer k. For each of the following assertions, give a proof or a counterexample:

(a) $|n_i| = 1$ for some i.

(b) If further all n_i are positive, then

$$\{n_1, n_2, \ldots, n_s\} = \{1, 2, \ldots, s\} \,.$$

B-5. For each $x > e^e$ define a sequence $S_x = u_0, u_1, u_2, \ldots$ recursively as follows: $u_0 = e$, while for $n \geq 0$, u_{n+1} is the logarithm of x to the base u_n. Prove that S_x converges to a number $g(x)$ and that the function g defined in this way is continuous for $x > e^e$.

B-6. Let $K(x,y,z)$ denote the area of a triangle whose sides have lengths x, y, and z. For any two triangles with sides a,b,c and a',b',c', respectively, prove that

$$\sqrt{K(a,b,c)} + \sqrt{K(a',b',c')} \leq \sqrt{K(a + a', b + b', c + c')}$$

and determine the cases of equality.

THE DOLCIANI MATHEMATICAL EXPOSITION SERIES.

All seven of the books in this widely acclaimed series contain fascinating problems that have ingenious solutions and/or involve intriguing results. These books will challenge the advanced mathematician as well as stimulate the mathematically talented high school student.

The seven volumes available in this series are:

MATHEMATICAL GEMS I, by Ross Honsberger. xi + 176 pages, Hardbound.
List: $16.50. MAA Member: $12.00.

MATHEMATICAL GEMS II, by Ross Honsberger. ix + 182 pages. Hardbound.
List: $16.50. MAA Member: $12.00.

MATHEMATICAL MORSELS, by Ross Honsberger. xii + 249 pages. Hardbound.
List: $21.00. MAA Member: $16.00.

MATHEMATICAL PLUMS, edited by Ross Honsberger. Articles by R.P. Boas, G. D. Chakerian, H. L. Dorwart, D. T. Finkbeiner, Ross Honsberger, K. R. Rebman, and S. K. Stein. ix + 182 pages. Hardbound.
List: $18.00. MAA Member: $13.50.

GREAT MOMENTS IN MATHEMATICS BEFORE 1650, by Howard Eves. xiv + 269 pages. Hardbound.
List: $22.00. MAA Member: $17.00.

MAXIMA AND MINIMA WITHOUT CALCULUS, by Ivan Niven, xv + 323 pages. Hardbound
List: $24.50. MAA Member: $18.50.

GREAT MOMENTS IN MATHEMATICS AFTER 1650, by Howard Eves, xii + 263 pages. Hardbound.
List: $23.50. MAA Member: $17.50.

***SPECIAL TWO-VOLUME* GREAT MOMENTS PACKAGE**—"Great Moments... Before and After 1650
List: $39.00. MAA Member: $29.50.

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

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 56, NO. 2, MARCH 1983