Vol. 77, No. 4, October 2004

MATHEMATICS MAGAZINE

- Where the Camera Was
- Tic-Tac-Toe on a Finite Plane
- Designs, Geometry, and a Golfer's Dilemma
- Arithmetic Progressions with Three Parts in Prescribed Ratio and a Challenge of Fermat

An Official Publication of The MATHEMATICAL ASSOCIATION OF AMERICA

EDITORIAL POLICY

Mathematics Magazine aims to provide lively and appealing mathematical exposition. The *Magazine* is not a research journal, so the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Manuscripts on history are especially welcome, as are those showing relationships among various branches of mathematics and between mathematics and other disciplines.

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 74, pp. 75–76, and is available from the Editor or at www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Submit new manuscripts to Frank A. Farris, Editor, Mathematics Magazine, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053-0373. Manuscripts should be laser printed, with wide line spacing, and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should mail three copies and keep one copy. In addition, authors should supply the full five-symbol 2000 Mathematics Subject Classification number, as described in *Mathematical Reviews*.

Cover image: see p. 274.

AUTHORS

Katie Byers graduated from Smith College with a degree in mathematics in 2002. She is currently teaching at Phillips Exeter Academy and hopes to earn a Ph.D. in either number theory or analysis in the next few years. She is an avid rower, and enjoys reading murder mysteries and linguistic theory.

Jim Henle is a set theorist whose usual research interests concern very large infinite sets. He was an undergraduate at Dartmouth College, and earned his doctorate from M.I.T. He wrote/co-wrote *Infinitesimal Calculus, An Outline of Set Theory* and *Sweet Reason*. His latest co-product is a calculus text that will appear early in 2005. In addition to interests in philosophy, education, and art, he is an earnest musician and an aggressive cook.

Maureen T. Carroll is an associate professor at the University of Scranton. She and her coauthor first started their collaborations when they were both graduate students at Lehigh University. Although her dissertation field was functional analysis, she has also published papers in voting theory and game theory. She was a Project NExT fellow (green dot) and participated in the Institute in the History of Mathematics.

Steven T. Dougherty received his doctorate from Lehigh University and is now a professor of mathematics at the University of Scranton. He has written papers in coding theory, number theory, and combinatorics with eighteen different coauthors and has lectured in six countries.

Keith E. Mellinger earned his Ph.D. in finite geometry at the University of Delaware. After graduate school he spent 2 years as a VIGRE postdoc at the University of Illinois at Chicago, studying coding theory and graph theory. He currently lives in Fredericksburg, VA, where he is an assistant professor at Mary Washington College. Keith's interests include many areas of discrete mathematics and, more importantly, connections between these areas. As they say, this article is based on a true story with minor modifications made for dramatic purposes. When not working on his family's interesting math problems, Keith plays the guitar and mandolin, and performs regularly both with bands and as a solo act.

Kenneth Fogarty is Professor Emeritus of Mathematics and Computer Science at Bronx Community College, CUNY. He resides in the Catskill Mountains region of New York state, where he enjoys canoeing and gardening in spring, summer, and fall, and shovels snow in the winter. His interest as a problemist is to pose problems that appeal to a broad audience. His other interests include wildlife conservation and habitat restoration.

Cormac O'Sullivan grew up in Ireland, received his B.A. from Trinity College Dublin and Ph.D. from Columbia University in 1998. After two years at the University of Maryland he joined the faculty at Bronx Community College. There he is pleased to be a mentor for the New York City Alliance for Minority Participation in Science, Engineering and Mathematics (NYC AMP). His mathematical interests are mostly in number theory, especially automorphic forms. Outside of that, he loves to explore New York City and hike around Ireland in the summers.

Vol. 77, No. 4, October 2004

MATHEMATICS MAGAZINE

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, bimonthly except July/August.
The annual subscription price for *MATHEMATICS MAGAZINE* to an individual member of the Association is $131. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% discount; and new members receive a 20% dues discount for the first two years of membership.)

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

Advertising correspondence should be addressed to Frank Peterson (*FPetersonj@aol.com*), Advertising Manager, the Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

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

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

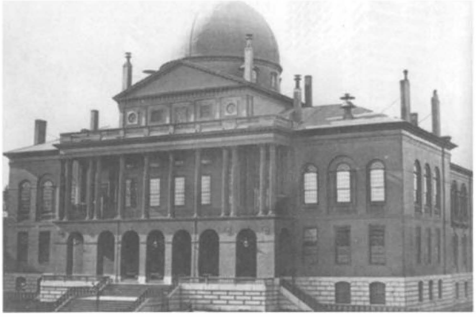
Printed in the United States of America

Where the Camera Was

KATHERINE McL. BYERS
JAMES M. HENLE
Smith College
Northampton, MA 01063
jhenle@math.smith.edu

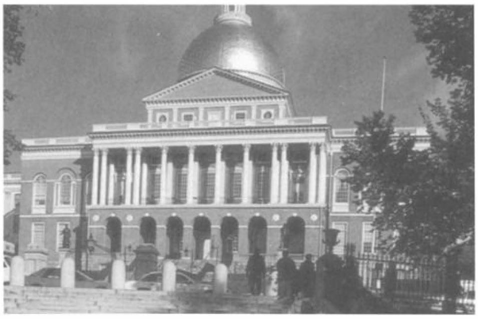
How many times have you seen something like this?

Then



Sources: Courtesy of The Bostonian Society/Old State House

Now



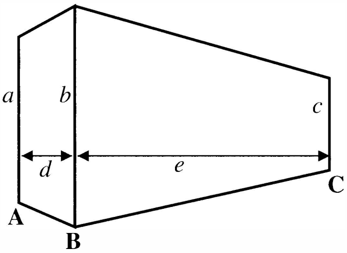
Simon Clay/Chrysalis Images

On the left is a picture of the Massachusetts Statehouse in Boston, taken about 1860. On the right is a picture taken in 1999. They appear in *Boston Then and Now* [**3**] and are meant to show us how the building and its setting have changed, but the effect is diminished because the camera was not in the same place for both photographs. How hard is it to determine the exact location of the photographer from information in a photograph?

The problem of understanding the relative positions of image and object is actively studied by computer scientists. In Kanatani [**2**], it is part of "computational projective geometry." The specific task of locating the camera from the photograph is called "camera calibration." In Kanatani's book the process is quite involved and technical. In a mathematical paper published later, Eggar [**1**] tackles the same problem. He proves that the task can be done, but the technique is similarly complex and the paper does not derive a practical method or formula.

In this paper, we present a method and a formula for locating the position of the photographer. Our basic result is the following:

PROPOSITION. *If a picture of a rectangular solid taken by a vertically-held pinhole camera has measurements (on the photograph) of a, b, c, d, and e,*



then the camera was positioned

$$\frac{dc}{d(b-c)+e(b-a)}\overline{\mathbf{BC}}$$

to the left of **B** *in the direction from* **C** *to* **B** *and*

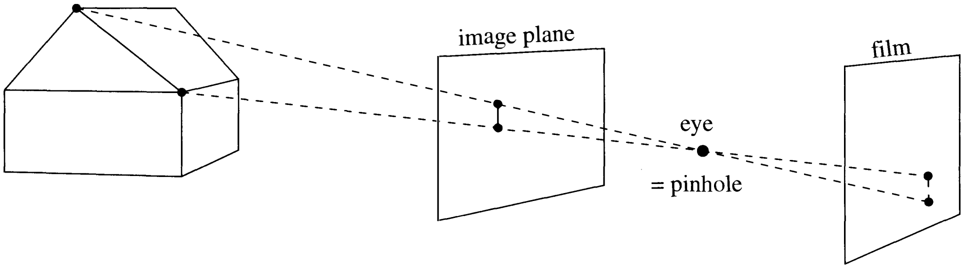
$$\frac{ae}{d(b-c)+e(b-a)}\overline{\mathbf{AB}}$$

in front of point **B**, *where* $\overline{\mathbf{BC}}$ *and* $\overline{\mathbf{AB}}$ *are on-site measurements.*

The proof is based on high-school plane geometry and the basic principles of projective geometry taught in a beginning drawing class.

Background

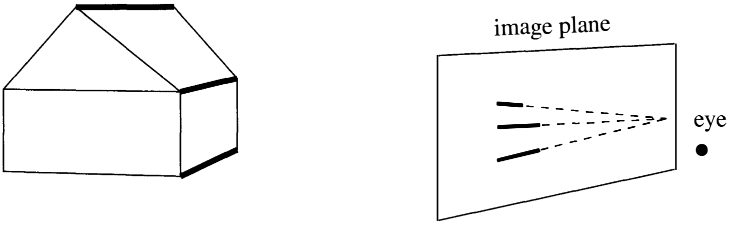
Our assumption is that the camera is a pinhole camera with the film in a vertical plane (plane perpendicular to the ground). Under these circumstances, the image on the film is the same as if we projected the three-dimensional world onto a plane, what we'll call the "image plane," using straight lines to the viewer's eye.



The only difference is that with a pinhole camera, the image appears on the film upside down.

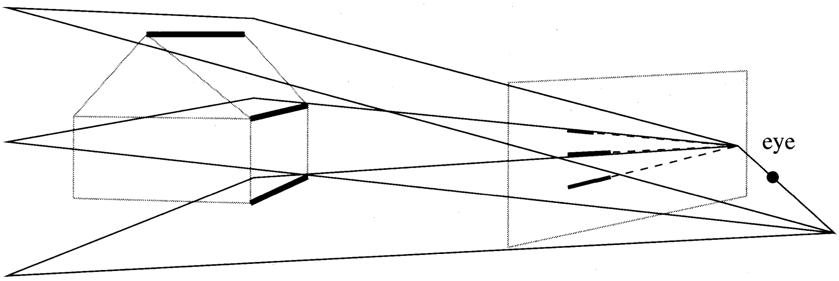
We'll need a few elementary facts about this projection:

(A) *The images of lines that are parallel to the ground and to one another, but not parallel to the image plane, meet at a single point in the image plane.*



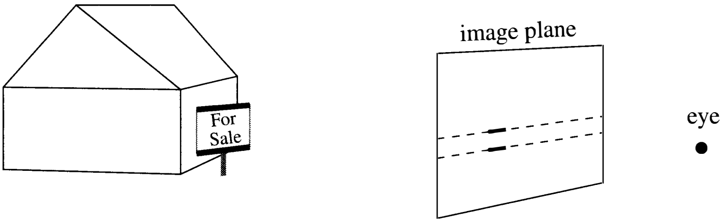
This point is called the **vanishing point** of the collection of parallel lines.

Imagine a collection of planes, each passing through the eye and one of the parallel lines. Then the planes intersect in a line that meets the image plane at the vanishing point.



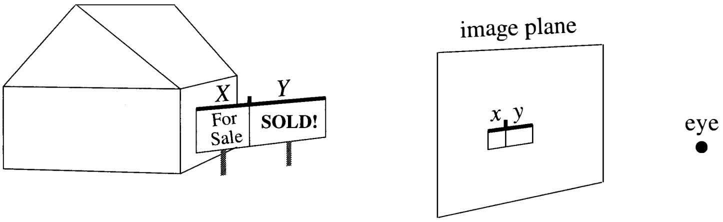
All such vanishing points lie on a single horizontal line called the **horizon line**.

(B) *Lines in the real world that are parallel to each other and also parallel to the image plane are parallel when projected onto the image plane.*



From this it follows that real horizontal lines are projected to horizontal lines.

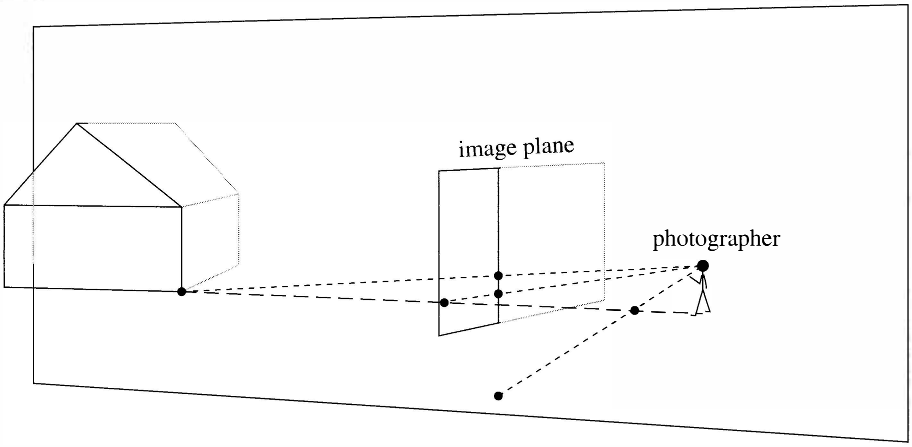
(C) *Also, ratios along lines parallel to the image plane are preserved when projected to the image plane. In the diagram below, this means that $X/Y = x/y$.*



Finally,

(D) *Lines on the ground connecting an object to the photographer appear as vertical lines on the image plane.*

Again, imagine a plane containing the eye of the photographer and the line to the photographer.



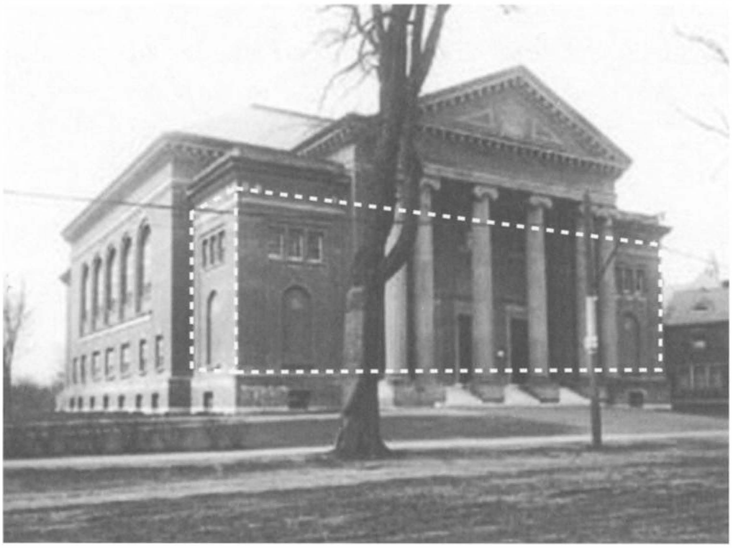
That plane is vertical and intersects the image plane in a vertical line.

A converse of **(D)** is also true: lines in the ground plane whose images are vertical connect to the photographer.

Our method

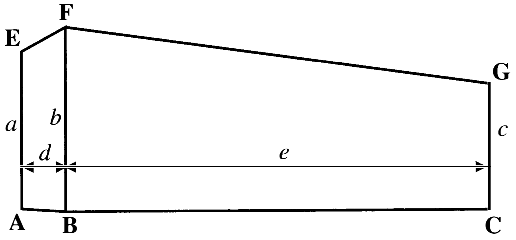
Given the tools above, we present a simple method for determining the location of the photographer.

We start with a photograph of John M. Greene Hall at Smith College, taken around 1935 by Edgar Scott. Since the building is a complex solid, we pick a rectangular solid on it whose corners are easy to locate.

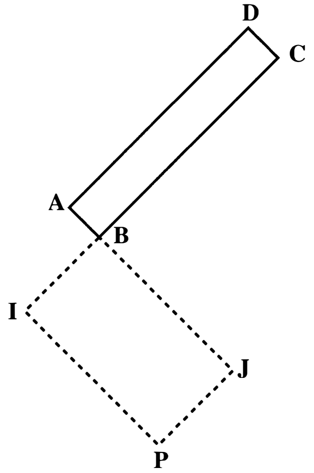


Source: Historic Northampton, Northampton, Massachusetts

We'll call this outline the *schematic picture*.



The schematic corresponds to the aerial view below, where **BC** is the front of the building and **P** is the location of the photographer.

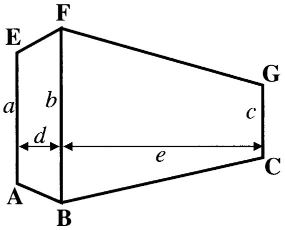


Our goal is to compute the distances $\overline{\mathbf{IB}}$ and $\overline{\mathbf{JB}}$. We'll compute $\overline{\mathbf{IB}}$—the computation of $\overline{\mathbf{JB}}$ can be done symmetrically. Our procedure is to express

$$\frac{\overline{\mathbf{IB}}}{\overline{\mathbf{BC}}}$$

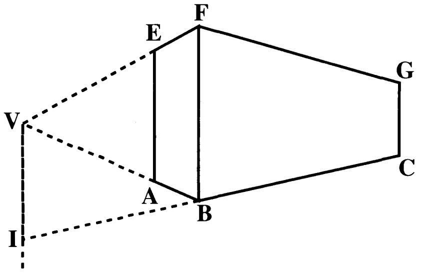
in terms of the five measurements a, b, c, d, and e in the image plane. Assuming we can measure $\overline{\mathbf{BC}}$ on site, we can then multiply this times the ratio to find $\overline{\mathbf{IB}}$.

To make the proof easier to view, we will show our work on a schematic with sharper angles:

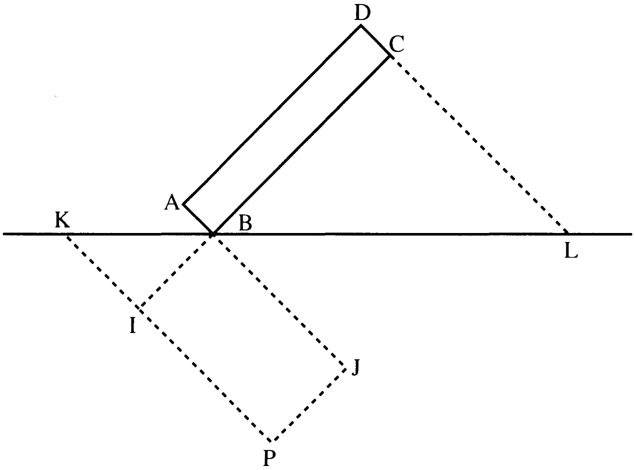


We begin by extending **EF** and **AB** in the schematic picture to determine the location of the left vanishing point, **V**.

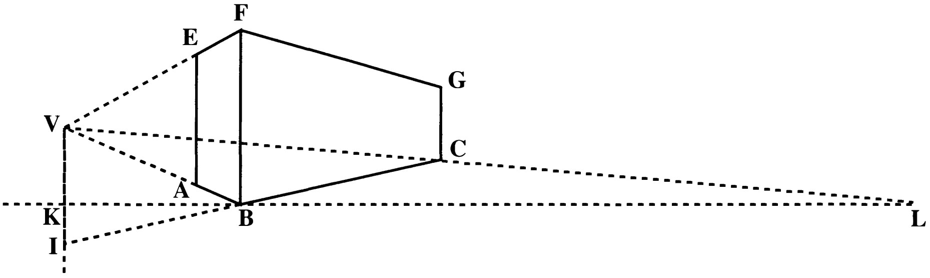
Next, notice that **PI** in the aerial view is parallel to **AB**, hence by Fact **(A)**, in the schematic picture it passes through **V**. Also, since it is a line to the photographer, by Fact **(D)** it is vertical in the schematic picture. Thus point **I** is the intersection of this vertical with the extension of **BC**.



Now we add a horizontal line through **B** parallel to the image plane and extend **PI** and **DC** to meet it. In the aerial view, it looks like:



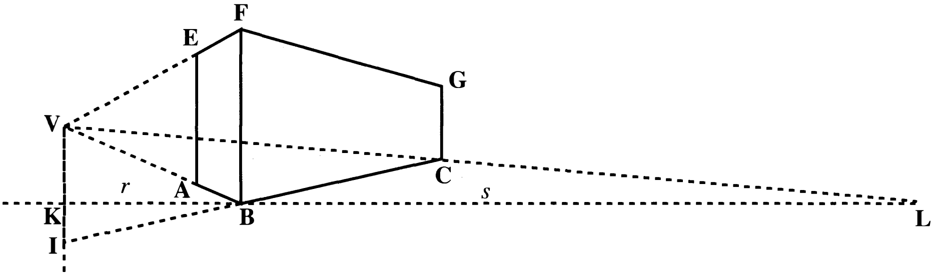
By Fact **(B)**, this line is also horizontal in the schematic. The aerial view line **CL** is parallel to **AB** and **PI**, so it too passes through **V**.



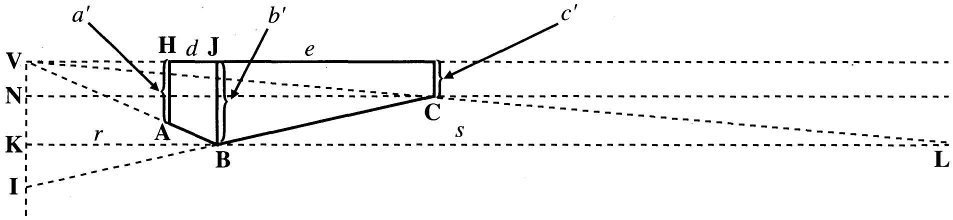
From $\triangle$**KIB** $\sim$ $\triangle$**LCB** in the aerial view we have

$$\frac{\overline{\mathbf{IB}}}{\overline{\mathbf{BC}}}=\frac{\overline{\mathbf{KB}}}{\overline{\mathbf{BL}}}.$$

From Fact **(C)**, this proportion is equal to the ratio of image plane distances r/s.



To find r/s, we add two more horizontal lines, **CN** and the horizon line **VH**, then focus on the lower half of the resulting figure.



From $\triangle$**VLK** $\sim$ $\triangle$**VCN** we have

$$\frac{r+s}{b'} = \frac{r+e}{c'}, \quad \text{from which we can derive:} \quad \frac{r}{s} = \frac{c'r}{b'r + b'e - c'r}.$$

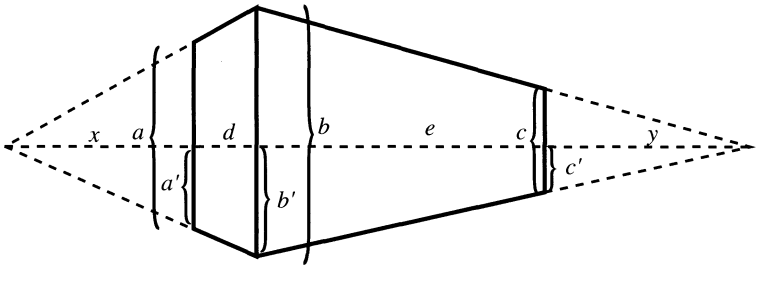
From $\triangle$**VJB** $\sim$ $\triangle$**VHA** we have

$$\frac{r}{b'} = \frac{r-d}{a'}, \quad \text{from which we can derive:} \quad r = \frac{b'd}{b'-a'}.$$

These together give us

$$\frac{r}{s} = \frac{c'\frac{b'd}{b'-a'}}{b'\frac{b'd}{b'-a'} + b'e - c'\frac{b'd}{b'-a'}} = \frac{c'd}{b'd + b'e - ea' - c'd}.$$

We promised to express this ratio in terms of a, b, c, d, and e. We can accomplish that by one more application of similar triangles: We have



$$\frac{a'}{b'} = \frac{x}{x+d} = \frac{a}{b}, \quad \text{and} \quad \frac{c'}{b'} = \frac{y}{y+e} = \frac{c}{b},$$

and so

$$\frac{a}{a'} = \frac{b}{b'} = \frac{c}{c'}$$

giving us

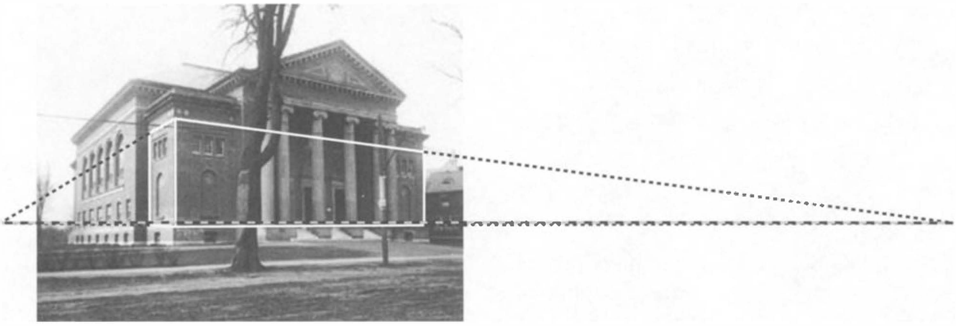
$$\frac{\overline{\mathbf{IB}}}{\overline{\mathbf{BC}}} = \frac{r}{s} = \frac{\frac{b}{b'}c'd}{\frac{b}{b'}b'd + \frac{b}{b'}b'e - e\frac{b}{b'}a' - \frac{b}{b'}c'd} = \frac{dc}{d(b-c) + e(b-a)}.$$

The corresponding formula for $\overline{\mathbf{BJ}}/\overline{\mathbf{AB}}$ can be found symmetrically:

$$\frac{\overline{\mathbf{BJ}}}{\overline{\mathbf{AB}}} = \frac{ae}{d(b-c) + e(b-a)}.$$

This completes the proof of the proposition. ■

The last step in locating the position of the camera is finding its height. This is accomplished in a primitive way by noting where the horizon line cuts across the picture. The height of the camera is the height of this line as it appears against the building in the picture.

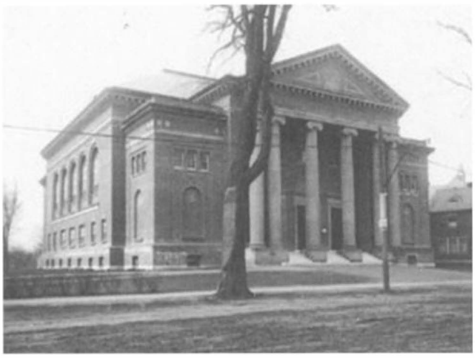


Source: Historic Northampton, Northampton, Massachusetts

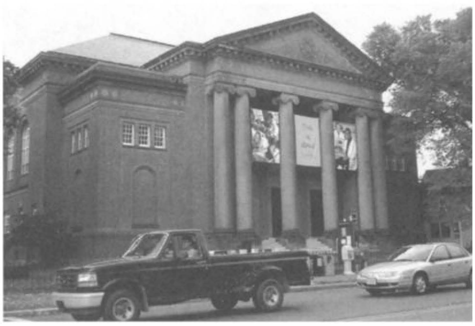
Conclusion

The close agreement of the two pictures illustrates the proposition.

Then



Now



Source: Historic Northampton, Northampton, Massachusetts

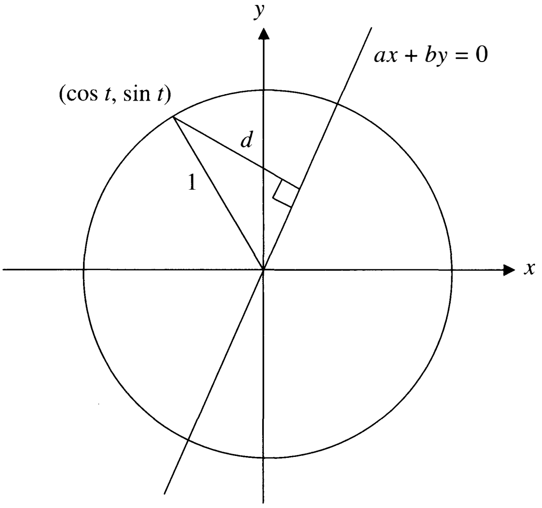
There are problems, though, in applying the proposition. It may be difficult to find an appropriate part of a building to analyze. It can be difficult to measure the building. It can be difficult to measure the photograph. Finally, locating the spot computed by the proposition, is not easy without equipment.

Considering these problems, the close agreement of the pictures of John M. Greene Hall might be considered good luck. We used a high-resolution scan on the archive photograph—b was measured at 470 pixels. Even so, if b were measured just one pixel less, the computed location of the photographer changes by almost two feet (because of the strategic location of b in the denominator of the formula).

REFERENCES

1. M. H. Eggar, Pinhole cameras, perspective, and projective geometry, *Amer. Math. Monthly* **105**:7 (1998), 618–630.
2. Kenichi Kanatani, *Geometric Computation for Machine Vision*, Clarendon Press, Oxford, 1993.
3. Elizabeth McNulty, *Boston Then and Now*, Thunder Bay Press, 1999.

Proof Without Words: Extrema of the Function $a\cos t + b\sin t$



$$d \le 1 \Rightarrow |a\cos t + b\sin t|/\sqrt{a^2+b^2} \le 1$$

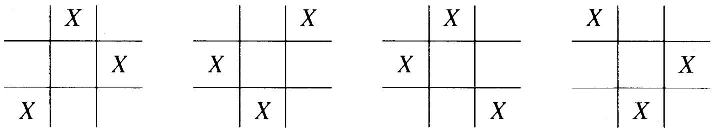
$$-\sqrt{a^2+b^2} \le a\cos t + b\sin t \le \sqrt{a^2+b^2}$$

—M. HASSANI, M. BAYAT, AND H. TEIMOORI
INSTITUTE FOR ADVANCED STUDIES IN BASIC SCIENCES,
P. O. BOX 45195-159,
GAVA ZANG, ZANJAN 45195, IRAN
Hassani@iasbs.ac.ir
Bayat@iasbs.ac.ir
Teimoori@iasbs.ac.ir

Tic-Tac-Toe on a Finite Plane

MAUREEN T. CARROLL
STEVEN T. DOUGHERTY
University of Scranton
Scranton, PA 18510
carrollm1@uofs.edu
doughertys1@uofs.edu

Everyone knows how to play tic-tac-toe. On an $n \times n$ board, if a player places n of her marks either horizontally, vertically, or diagonally before her opponent can do the same, then she wins the game. What if we keep the rules of the game the same but increase the number of ways to win? For simplicity, any configuration of n marks that produces a win, regardless of whether or not it appears straight, will be called a winning line. For example, we will add the four winning lines shown in FIGURE 1 when playing on the 3×3 board.



	X	
		X
X		

		X
X		
	X	

	X	
X		
		X

X		
		X
	X	

Figure 1 New winning lines for 3×3 tic-tac-toe

This brings the total number of winning lines on this board to twelve. Why did we decide to add these particular lines? If you know the rudimentaries of finite geometry, you can see that the winning lines are prescribed by the geometry of a finite affine plane. Otherwise, for now you should just notice that every new line contains exactly one mark in each row and each column. You should also notice that these new lines make it more difficult to identify a win here than in the standard game. As you will see, the reason for this complexity is that lines in an affine plane need not appear straight. With this new twist, the game that grew tiresome for us as children is transformed into an interesting, geometrically motivated game.

The geometric intuition required to understand finite planes often proves elusive, as our Euclidean-trained minds have preconceived notions of lines and points. The new version of tic-tac-toe helps to develop this intuition. Moreover, this game relates geometric concepts to game-theoretic concepts as the natural question of winning strategies arises. Since more winning lines mean more possible ways to win, one might think that it would be easier to force a win in this new game. Not only is the answer to this question nonintuitive, but the difficulty encountered in providing an answer for the 4×4 board is surprising.

First, we review Latin squares and affine planes, as well as the relationship between them, in order to find the new winning lines. Once you can identify the winning lines, you are ready to play tic-tac-toe on the affine plane. Since projective planes are a natural extension of affine planes, you will also learn to play tic-tac-toe on these planes. You may recall that in the 3×3 version of tic-tac-toe we played as children, one quickly learns that there is no advantage to being the first player since the game between two skilled players always ends in a draw. While this is the case on many finite planes, we will show that there *are* planes where the first player holds the advantage. In the event that you are the *second* player on a plane where a forced draw is possible, we provide a computational method that guarantees a draw. We will also show simple configurations of points that produce a draw with very few points.

Squares and planes

Consider the 36 officer problem: There are 36 officers, each with one of six rank designations and one of six regiment designations. Can the 36 officers be arranged into six rows and six columns so that each rank and regiment is represented in each row and each column? Leonhard Euler showed this could be done for 9 and 25 officers (try it!), but conjectured correctly that it could not be done for 36 officers. In an attempt to solve this problem he introduced Latin squares [**10**]. A Latin square of order n is an $n \times n$ matrix with entries from $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$, where each number occurs exactly once in each row and each column. Examples of Latin squares of orders 2, 3, and 4 are given in FIGURE 2.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 & 3 \\ 2 & 3 & 0 & 1 \\ 3 & 2 & 1 & 0 \\ 1 & 0 & 3 & 2 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \end{bmatrix}$$

Figure 2 Latin squares of orders 2, 3, and 4

Since it is natural to explain the game of tic-tac-toe on a finite plane by the connection between planes and these squares, we begin with an explanation of Latin squares, affine planes, and the relationship between them. The material presented in this section can be found in any text on affine and projective planes [**3, 17**]. Readers familiar with these concepts may wish to proceed to the next section.

Latin squares $A = [a_{ij}]$ and $B = [b_{ij}]$ are orthogonal if and only if $C = [c_{ij}]$, whose entries are the ordered pairs $c_{ij} = (a_{ij}, b_{ij})$, contains all n^2 possible ordered pairs of $\mathbb{Z}_n \times \mathbb{Z}_n$. A collection of Latin squares is mutually orthogonal (MOLS) if and only if each pair is orthogonal. (The Maple command `MOLS(p,m,n)` produces n MOLS of order p^m when p is prime and $n < p^m$.) In the examples above, the two Latin squares of order 3 are orthogonal, and the three of order 4 are MOLS. Euler's 36 officer problem asks if it is possible to find a pair of orthogonal Latin squares of order 6, one representing the ranks of the 36 officers and the other representing the regiments. As illustrated by the first Latin square of order 3 in FIGURE 2, you can easily produce one Latin square of order 6 by continually shifting the elements of your first row to the right by one position and wrapping the leftover elements to the beginning. The proof of the 36 officer problem shows that you cannot produce a second Latin square orthogonal to the first. (Try it!) Exhaustive solutions [**19**] to this problem, as well as more sophisticated ones [**8, 18**], can be found in the literature. (Laywine and Mullen [**14**] offer many interesting questions concerning Latin squares.)

The Euclidean plane is an example of an affine plane, and the axioms of affine planes are merely a subset of those from Euclidean geometry. Specifically, an affine plane is a nonempty set of points, P, and a nonempty collection of subsets of P (called lines), L, that satisfy the following three axioms:

(1) through any two distinct points there exists a unique line;

(2) if p is a point, ℓ is a line, and p is not on line ℓ, then there exists a unique line, m, that passes through p and is parallel to ℓ, that is, $p \in m$ and $\ell \cap m = \emptyset$;

(3) there are at least two points on each line, and there are at least two lines.

When p is a point on line ℓ, we say that p is incident with ℓ. The Cartesian plane, with points and lines defined as usual, is the example we typically envision when reading this definition. It is an example of an infinite affine plane.

Finite affine planes are those with a finite set of points. There is no finite affine plane where P contains exactly one, two, or three points. (Why not? What axiom(s) of affine planes would such situations violate?) The smallest finite affine plane can be given as $P = \{p, q, r, s\}$ and $L = \{\{p, q\}, \{p, r\}, \{p, s\}, \{q, r\}, \{q, s\}, \{r, s\}\}$, which is represented by either of the graphs in FIGURE 3. Notice that an intersection of line segments does not necessarily indicate the existence of a point in P.

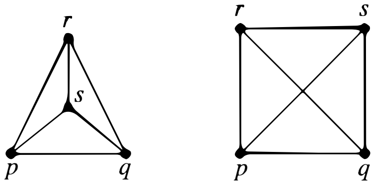


Figure 3 Two graphical representations of the affine plane of order 2

Using the given axioms, we invite the reader to reproduce the following elementary results: On a finite affine plane, each line must contain the same number of points and each point is incident with the same number of lines. The number of points on each line is called the order of the plane. This is why the diagrams in FIGURE 3 are described as the affine plane of order 2. In general, an affine plane of order n has n points on every line, and each point is incident with $n + 1$ lines. For any such plane, $|P| = n^2$ and $|L| = n^2 + n$. Two lines are parallel if and only if they have no common points, and parallelism is an equivalence relation on the set of lines. A parallel class consists of a line and all the lines parallel to it. An affine plane of order n has $n + 1$ parallel classes, each containing n lines. As another example, FIGURE 4 shows the affine plane of order 3, where $P = \{a, b, c, d, e, f, g, h, i\}$ and $L = \{\{a, b, c\}, \{d, e, f\}, \{g, h, i\}, \{a, d, g\}, \{b, e, h\}, \{c, f, i\}, \{a, e, i\}, \{c, e, g\}, \{a, h, f\}, \{g, b, f\}, \{i, b, d\}, \{c, h, d\}\}$. You can see that each line has three points, each point is incident with four lines, $|P| = 9$, and $|L| = 12$. The lines $\{c, e, g\}$, $\{a, h, f\}$, and $\{i, b, d\}$ are parallel and, therefore, form one of the four parallel classes.

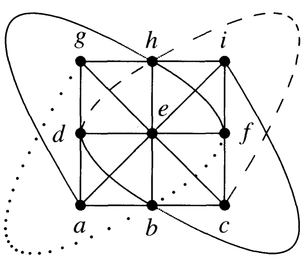


Figure 4 Affine plane of order 3

Although we have seen affine planes of orders 2 and 3, for some orders there is no such plane. In fact, determining which orders of affine planes exist is exceptionally difficult, and remains a largely open problem. It *is* well known that there are affine planes of order p^k where p is prime and $k \in \mathbb{Z}^+$. (You can read about these in Mellinger's article in this issue of the MAGAZINE.) This tells us, for example, that there are affine planes of orders 2, 3, 4, 5, 7, 8, and 9. How about 6 and 10? We can answer one of these using the following connection between affine planes and Latin squares: Bose [7] showed that an affine plane of order n exists if and only if there exist $n-1$ MOLS of order n. Using this result, we see that there can be no affine plane of order 6 since the solution to the 36 officer problem shows that there is no *pair* of orthogonal Latin squares of order 6. The proof of the nonexistence of the plane of order 10 is much more difficult, requiring a great deal of mathematics and an enormous computation to finish the proof. (Lam [13] gives an historical account.) It is not known whether an affine plane of order 12 exists. In fact, it is unknown whether there are any affine planes that do not have prime-power order. Planes of some composite orders, however, are known *not* to exist (see the Bruck-Ryser Theorem [3]).

This connection between affine planes of order n and the $n-1$ MOLS of order n can be used to find the lines of the plane quite easily. After arranging the n^2 points of a finite affine plane in an $n \times n$ grid, we will first identify its $n+1$ parallel classes, which in turn reveals all of the lines. The n horizontal lines form one parallel class, and the n vertical lines form another. Each of the remaining $n-1$ parallel classes corresponds to one of the $n-1$ MOLS as follows: the ith line in any parallel class is formed by the positions of symbol i in the corresponding Latin square. (Here, $i = 0, 1, \dots, n-1$.) For example, using FIGURE 4 and the two orthogonal 3×3 Latin squares in FIGURE 2, we see that the four parallel classes for the affine plane of order 3 are

(i) the horizontal lines $\{\{a, b, c\}, \{d, e, f\}, \{g, h, i\}\}$,

(ii) the vertical lines $\{\{a, d, g\}, \{b, e, h\}, \{c, f, i\}\}$,

(iii) the lines indicated by the first Latin square $\{\{c, e, g\}, \{a, h, f\}, \{i, b, d\}\}$, and

(iv) the lines indicated by the second Latin square $\{\{g, b, f\}, \{c, h, d\}, \{a, e, i\}\}$.

At this point you might notice that the four lines that do not appear to be straight correspond precisely to the winning lines we added to the 3×3 tic-tac-toe board, as shown in FIGURE 1.

There is one other type of plane on which we will play tic-tac-toe, namely, a *finite projective plane*. A projective plane is easily constructed from an affine plane of order n by adding $n+1$ points (the points at infinity) and one line (the line at infinity). The points are added in this way: Each point at infinity must be incident with the n lines of a unique parallel class. (Now you see that $n+1$ points must be added since there are $n+1$ parallel classes on the affine plane of order n.) The line at infinity, ℓ_∞, simply consists of the $n+1$ points at infinity. For example, the projective plane of order 2 can be constructed from the affine plane of order 2 given in FIGURE 3 by adding $\ell_\infty = \{a, b, c\}$, as shown in both of the graphs in FIGURE 5. Here we see point a is added to the parallel lines $\{r, p\}$ and $\{s, q\}$, b is added to the parallel lines $\{r, s\}$ and $\{p, q\}$, and c is added to the parallel lines $\{q, r\}$ and $\{p, s\}$.

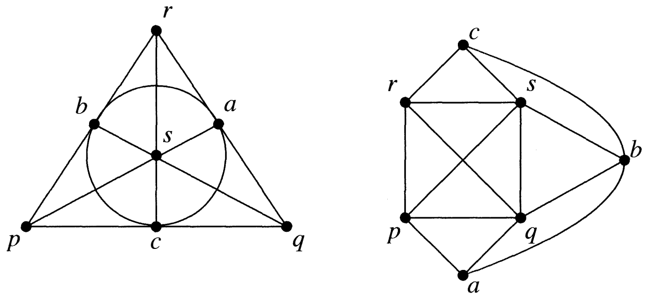


Figure 5 Two graphical representations of the projective plane of order 2

Of course, we could have discussed projective planes before affine planes. By definition, a projective plane is a nonempty set of points, P, and a nonempty set of lines, L, that satisfy the following three axioms: (1) any two distinct lines meet in a unique point; (2) any two distinct points have a unique line through them; (3) there are at least three points on each line and there are at least two lines. We invite the reader to reproduce the following elementary results: On a finite projective plane, each line must contain the same number of points. In particular, a projective plane of order n has $n+1$ points on each line. For any such plane, $|P| = |L| = n^2 + n + 1$. For example, in FIGURE 5 we can identify the seven lines of the plane of order 2 as $\{r, p, a\}$, $\{s, q, a\}$, $\{r, s, b\}$, $\{p, q, b\}$, $\{q, r, c\}$, $\{p, s, c\}$, and $\{a, b, c\}$. Since the number of points is not a perfect square, you should notice that we will not be playing tic-tac-toe on an $n \times n$ grid for these planes. Lastly, just as we were able to construct a projective plane from an affine plane, we can do the opposite. Starting with a projective plane of order n, removing any line and all of the points with which it is incident forms an affine plane of order n.

Throughout this work the word plane will refer to either an affine or projective plane. Affine planes of order n will be represented as π_n, and projective planes as Π_n. All statements about uniqueness are always understood to mean up to isomorphism.

The game

A zero-sum game is a game where one player's loss is a gain for the other player(s). The standard game of tic-tac-toe is a two-player, zero-sum game on a 3×3 board where players alternately mark one open cell with an X or an O. For simplicity, we will refer to player X as Xeno, player O as Ophelia, and assume that Xeno always makes the first move. A player wins by being the first to place three matching marks on a line. If a game is complete and no player has won, the game is a draw. Tic-tac-toe is an example of a game of perfect information since each choice made by each player is known by the other player. Poker is not such a game since players do not reveal their cards.

A *strategy* is an algorithm that directs the next move for a player based on the current state of the board. A winning strategy for Xeno, for example, is a strategy that is guaranteed to produce a win for him. Although there is a best way to play standard tic-tac-toe, there is no winning strategy since each player can guarantee that the other cannot win. In this case, we say that both players have a drawing strategy, that is, an algorithm that leads to a draw. The assumption that both players are knowledgeable and play correctly is a standard game-theoretic assumption called the principle of rationality, that is, at each move, each player will make a choice leading to a state with the greatest utility for that player.

We give the following definition for tic-tac-toe on a plane of order n. Xeno and Ophelia alternately place their marks on any point of the plane that has not already been labelled. The first player to claim all of the points on a line wins the game. The game is a draw if all points are claimed and neither player has completed a line. In order to show the order of play, we will denote Xeno's first move as X_1, his second move as X_2, and so on. Ophelia's moves are likewise designated. We shall refer to a game as an ordered pair $[(X_1, \ldots, X_s), (O_1, \ldots, O_r)]$ where $r = s - 1$ or $r = s$. A complete game is one that has resulted in a win or a draw.

As suggested in the previous section, when playing on π_n we will arrange the n^2 points in an $n \times n$ grid. In this way, the cells of the standard $n \times n$ tic-tac-toe grid have become the points of π_n. When Xeno marks an open cell in the grid with his X, he is essentially claiming a point on the affine plane. The $n^2 + n$ lines of π_n are found in this way: n lines are horizontal, n lines are vertical, and the remaining $n^2 - n$ lines are identified by consulting the $n - 1$ MOLS of order n. Remember, each of the MOLS of order n defines n lines (displayed as identical symbols). Since there are $n - 1$ MOLS, each defining n lines, we have our remaining $n(n - 1)$ lines. For example, the game shown on the left in FIGURE 6 is a win for Xeno on the affine plane of order 3 since $\{X_2, X_3, X_4\}$ forms a line, as can be verified by viewing FIGURE 4 or consulting the second Latin square of order 3 given in FIGURE 2. The game on the right is a win for Ophelia on the affine plane of order 4 since $\{O_1, O_3, O_6, O_7\}$ forms a line, as can be verified by consulting the second Latin square of order 4 in FIGURE 2.

X_1	X_4	
X_3	O_2	O_1
O_3		X_2

X_1	X_2	O_3	X_3
O_1	X_6	X_4	
X_7	O_6	O_2	O_5
O_4	X_5		O_7

Figure 6 Win for Xeno on π_3 and win for Ophelia on π_4

Of the many interesting graph-theoretic, game-theoretic, and combinatorial questions this game generates, we will first consider two fundamental questions.

Question 1: For which planes are there winning strategies?

Question 2: For which planes can play end in a draw?

The first question is essentially a game-theoretic question, whereas the second question is fundamentally a geometric question. As regards the first question, in game theory it is known that in a finite two-player game of perfect information, either one player has a winning strategy or both players can force a draw [**16**]. A "strategy-stealing" argument [**4, 5**] proved by Hales and Jewett [**11**] shows that in our case it is Xeno who has a winning strategy when such a strategy exists. To show this, assume that Ophelia has the winning strategy. Let Xeno make a random first move and thereafter follow the winning strategy of Ophelia. Specifically, Xeno plays as if he *were* Ophelia by pretending that his first move has not been made. If at any stage of the game he has already made the required move, then a random move can be made. Any necessary random moves, including the first, cannot harm him since he is merely claiming another point. This leads Xeno to a win, contradicting the assumption that Ophelia has the winning strategy. (Notice that this argument does not apply to Nim, for example, since a random move may cause the first player to lose.) Hence, in tic-tac-toe either Xeno has a winning strategy or both players have drawing strategies, in which case we say Ophelia can force a draw. If no draws exist, then Xeno is guaranteed to have a winning strategy. However, the existence of draws is not enough to guarantee that Ophelia can force a draw. We discuss the existence of winning strategies for all finite planes in the two sections that follow.

Regarding the second question, a draw is possible when there exists a set T of $\lceil |P|/2 \rceil$ points such that every line in the plane has points in T and points not in T, that is, no line has its points disjoint from T nor contained in T. We will determine the planes in which play can end in a draw in the following two sections. Of course, knowing that a draw exists does not explain how Ophelia can force the draw. To this end, we give a computational method guaranteed to produce a draw in the section on weight functions, and we describe simple configurations of draws in the last section of the paper.

Planes of small order

There is a unique affine plane of order 2; in it each line has two points, as represented in FIGURE 3. Xeno has a trivial winning strategy when playing tic-tac-toe on this plane. Namely, if X_1 and O_1 are chosen arbitrarily, then X_2 produces a win for Xeno with the line containing X_1 and X_2, regardless of its placement. Hence, Xeno wins merely by being the first player, and a draw is not possible since any two points form a line.

There is a unique projective plane of order 2; as represented in FIGURE 5, each line has three points. Xeno has a winning strategy when playing on this plane as well. Namely, if X_1 and O_1 are chosen arbitrarily, then he chooses X_2 to be any point not on the line containing X_1 and O_1. Since there is a line between any two points, O_2 must be placed on the line containing X_1 and X_2 (otherwise Xeno wins on his next move). He chooses X_3 to be the point on the line containing O_1 and O_2. Then Ophelia must block either the line containing X_1 and X_3 or the line containing X_2 and X_3 (it is a simple matter to see that Ophelia does not already have these lines blocked). Xeno wins on his next move when he completes the line that O_3 did not block. Even if the principle of rationality is violated and Xeno purposely chooses a point unwisely, a draw is not possible on Π_2. Any four points on Π_2, no three of which are collinear, form an object called a hyperoval, and the complement of this hyperoval is a line. Hence, there does not exist a set $T \subset P$ with $|T| = 4$ such that T and its complement intersect each line.

There is a unique affine plane of order 3. As shown in FIGURE 4, each line has three points. The winning strategy for Xeno on this plane is identical to the winning strategy on Π_2. (It is interesting to note that playing on this plane is the same as playing on a torus version of tic-tac-toe [**20**].) If the principle of rationality is violated then the game *could* end in a win for Ophelia, but a draw is impossible since there are no draws on π_3. To show this, assume that a draw *is* possible and let T be a set of five points that meets each line in L without containing any line completely. Let ℓ_1, ℓ_2, and ℓ_3 be the three lines of one of the parallel classes of π_3. Without loss of generality, assume T meets both ℓ_1 and ℓ_3 at two points and ℓ_2 at one point. Let T meet ℓ_1 at points x and y, and ℓ_2 at point z. The line between x and z intersects ℓ_3, say at p. The line between y and z also intersects ℓ_3, say at q. Notice that both of these lines go through z, and ℓ_3 is not parallel to either of these lines. Therefore, we have $p \neq q$ since no two lines intersect in more than one point. Since T intersects ℓ_3 in two points, if p is not in T then q must be in T. So, either line $\{x, z, p\}$ or $\{y, z, q\}$ is in T, which contradicts our assumption of the existence of a draw.

The following theorem summarizes our discussion of the analysis of play on the planes of small order.

THEOREM. *Xeno has a winning strategy on π_2, π_3, and Π_2, and no draw is possible on these planes.*

Weight functions and planes of larger order

When we venture beyond the planes of small order the complexity of the game increases dramatically. The additional points and lines generate a far greater number of possible moves for each player. This prevents an easy move-by-move analysis as we did in the previous section. This is where Erdős comes to our rescue. The two theorems that follow are special cases of a result of Erdős and Selfridge [**9**] that specifies conditions under which the second player can force a draw in many positional games. Our proofs are a modification of the proof of the Erdős and Selfridge theorem given by Lu [**15**].

To analyze the game on any plane of order n, we need a way to evaluate the state of the game at any point during play. It would be helpful to assign a number that in some way measures the utility of the state of the game for one of the players. To do this, we define functions that assign values to the state of the game when Ophelia is about to make her ith move. In order to choose the position for O_i from the unclaimed points remaining, she may first wish to consider which line has the best available point. Keep in mind that Ophelia forces a draw if she places one of her marks on every line, thereby blocking every possible winning line for Xeno. So, any line that Ophelia has already blocked can be removed from consideration. Of the unblocked lines remaining, it is most important for Ophelia to block lines with the largest number of Xeno's marks. If we define the value, or weight, of an unblocked line to be 2^{-u}, where u is the number of available points on that line, then the lines of greater weight are precisely those with more of Xeno's marks, and are therefore urgent for Ophelia to block. As Ophelia is about to make her ith move, the weight of the game is defined as the sum of the weights of the unblocked lines. The weight of an available point is the sum of the weights of any unblocked lines incident with the point. Lastly, the weight of a pair of available points on an unblocked line is the weight of the line through these points.

To give formulas to match these descriptions, we need some notation. Assume that the current state of play is $[(X_1, \ldots, X_i), (O_1, \ldots, O_{i-1})]$ and that L represents the set of lines. Let L_i be the collection of all lines not blocked by Ophelia at the ith move for Ophelia, with all of the points previously marked by Xeno deleted, that is, $L_i = \{\ell - \{X_1, \ldots, X_i\} \mid \ell \in L,\ \ell \cap \{O_1, \ldots, O_{i-1}\} = \emptyset\}$. So, L_i contains lines or subsets of lines that have not been blocked by Ophelia. We will let L_∞ denote the collection when no more moves can be made, that is, the game has ended in a win or a draw. We use ∞ rather than a particular number as the number of these collections depends on both the order of the plane and the progress of play. Let $P_i = P - \{X_1, \ldots, X_i, O_1, \ldots, O_{i-1}\}$, the set of points available to Ophelia at move O_i.

With this notation, the *weight of the game* is

$$w(L_i) = \sum_{s \in L_i} 2^{-|s|}.$$

For $p, q \in P_i$, the *weight of an available point* q and the *weight of an available pair* $\{p, q\}$ are

$$w(q \mid L_i) = \sum_{s \in L_i, q \in s} 2^{-|s|} \quad \text{and} \quad w(p, q \mid L_i) = 2^{-|s|}, \quad \text{where} \quad \{p, q\} \subseteq s \in L_i.$$

Let's compute examples of these various weights, using the game played on π_3 as shown on the left in FIGURE 6. Here, L_1 consists of eight lines of cardinality 3 and four partial lines of cardinality 2 (since X_1 has been removed), giving $w(L_1) = 4 \cdot 2^{-2} + 8 \cdot 2^{-3}$. For L_2 the state of the game is $[(X_1, X_2), (O_1)]$, and we eliminate the four lines through O_1 from consideration. Thus, L_2 consists of three lines of cardinality 3, four partial lines of cardinality 2, and one of cardinality 1, giving $w(L_2) = 2^{-1} + 4 \cdot 2^{-2} + 3 \cdot 2^{-3}$. Since there are four lines through any point on π_3, we see that $w(O_1 \mid L_1) = w(X_2 \mid L_1) = 2^{-2} + 3 \cdot 2^{-3}$. Also, $w(X_2, O_1 \mid L_1) = 2^{-3}$. Continuing this example, for L_3 the state of the game is $[(X_1, X_2, X_3), (O_1, O_2)]$, and we eliminate the seven lines through O_1 or O_2 from consideration. We have $w(L_3) = 2 \cdot 2^{-1} + 3 \cdot 2^{-2}$, $w(O_2 \mid L_2) = 2^{-1} + 2 \cdot 2^{-3}$, $w(X_3 \mid L_2) = 2 \cdot 2^{-2} + 2^{-3}$, and $w(X_3, O_2 \mid L_2) = 0$.

Consider the difference in weights between two successive states of the game, $w(L_i) - w(L_{i+1})$. The only change between L_i and L_{i+1} is that Ophelia's ith move and Xeno's $(i + 1)$st move have been made. So, the weights of any lines that do not contain O_i and X_{i+1} do not change and will therefore cancel each other out. With only the lines through these two points remaining, the weights of the lines through X_{i+1} must be subtracted from the weights of the lines through O_i in order to find $w(L_i) - w(L_{i+1})$. Since this eliminates the weight of the line that passes through both points, the weight of this line must be added back. Thus, it can be seen that

$$w(L_i) - w(L_{i+1}) = w(O_i \mid L_i) - w(X_{i+1} \mid L_i) + w(X_{i+1}, O_i \mid L_i). \qquad (1)$$

The examples given above can be used to demonstrate (1) when $i = 1$ and $i = 2$.

These weight functions enable us to check if we have a draw at any stage of play. First notice that if $\emptyset \in L_i$, then $w(L_i) \geq 2^{-0} = 1$, Xeno has completed a line and thus, has won. On the other hand, if $w(L_i) < 1$ then $\emptyset \notin L_i$, and Xeno has not completed a line. Also, notice that if $w(L_\infty) < 1$ then $\emptyset \notin L_\infty$ and there is a draw. Moreover, these weight functions provide strategies for Xeno and Ophelia that will help us determine the outcome of play on all planes of higher order. Namely, Xeno should minimize $w(L_i) - w(L_{i+1})$ in an attempt to keep the weight of L_j, at any stage j of the game, above 1, whereas Ophelia should maximize this difference in order to drag the overall weight below 1. Hence, by equation (1), Ophelia chooses O_i by maximizing $w(O_i \mid L_i)$, and Xeno chooses X_{i+1} by maximizing $w(X_{i+1} \mid L_i) - w(X_{i+1}, O_i \mid L_i)$. The power and utility of these weight functions is demonstrated in the proof of the following theorem, where the drawing strategy for Ophelia is specified for infinitely many projective planes.

DRAW THEOREM FOR Π_n. *Ophelia can force a draw on every projective plane of order n with $n \geq 3$.*

Proof. To prove that Ophelia can force a draw, we must produce an algorithm that prescribes Ophelia's move at any point in the game, and then show that this strategy leads to a draw. As noted above, if $w(L_\infty) < 1$ then Ophelia has forced a draw. This is equivalent to showing two conditions:

(i) There exists N, where $1 \leq N < \infty$, such that $w(L_N) < 1$ and
(ii) $w(L_{i+1}) \leq w(L_i)$ for all $i \geq N$.

Suppose that the current state of play is $[(X_1, \ldots, X_i), (O_1, \ldots, O_{i-1})]$, and Ophelia must make her ith move. Since the weight functions assign more weight to lines on which Xeno is closer to winning, Ophelia should choose a point of maximal weight. So, choose $O_i \in P_i$ such that $w(O_i \mid L_i) = \max\{w(q \mid L_i) : q \in P_i\}$. By the choice of O_i and (1), we see that the second condition is always satisfied since $w(O_i \mid L_i) \geq w(X_{i+1} \mid L_i)$.

For a projective plane of order n, L_1 consists of $n+1$ partial lines of cardinality n (once X_1 is removed) and $(n^2+n+1)-(n+1)$ lines of cardinality $n+1$. So, we have

$$w(L_1) = \sum_{i=1}^{n+1} 2^{-n} + \sum_{i=1}^{n^2} 2^{-(n+1)} = \frac{n^2+2n+2}{2^{n+1}}.$$

We see that $w(L_1) < 1$ when $n \geq 4$. Thus, Ophelia forces a draw on the projective planes of order $n \geq 4$ by choosing a point of maximum weight at every stage of the game.

For the projective plane of order 3, recall that there are 13 lines with 4 points on each line, and 4 lines through each point. We calculate $w(L_3)$ after providing the strategy for Ophelia's first two moves. Suppose X_1 and O_1 are placed arbitrarily. Xeno places X_2 anywhere. If O_1 is already on the line containing X_1 and X_2, then O_2 should not be placed on this line. If O_1 is not on the line containing X_1 and X_2, then O_2 should be placed on this line. In either case, the configuration before move X_3 is represented by FIGURE 7.

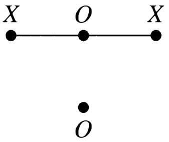


Figure 7 Configuration of $[(X_1, X_2), (O_1, O_2)]$ on Π_3

Xeno can place X_3 anywhere, leaving only four possible configurations of points, as represented in FIGURE 8. As long as $w(L_3) < 1$ in each case, then Ophelia has forced a draw.

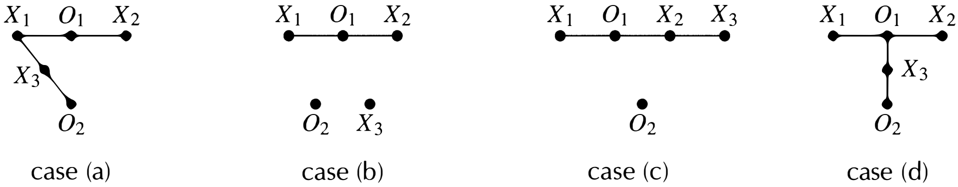


Figure 8 Possible configurations for $[(X_1, X_2, X_3), (O_1, O_2)]$ on Π_3

To calculate $w(L_3)$, in all four cases we start by eliminating the four lines containing O_1 and the remaining three lines containing O_2. Once these seven lines are eliminated from consideration, there are only six lines remaining to be included in the weight function.

Case (a): There are two partial lines through X_1 of cardinality 3. There is one partial line of cardinality 3 through X_2, and one partial line through X_2 and X_3 of cardinality 2. Through X_3 there is one remaining line of cardinality 3. Since only 12 out of 13 lines have been considered, there is one line of cardinality 4 remaining. This gives $w(L_3) = 2^{-2} + 4 \cdot 2^{-3} + 2^{-4} = 13/16$.

Case (b): There is one partial line through X_1 of cardinality 3, and one partial line through X_1 and X_3 of cardinality 2. The same holds for X_2. All lines through X_3 have been considered. Since only 11 of the 13 lines have been considered, there are two lines of cardinality 4 remaining. This gives $w(L_3) = 2 \cdot 2^{-2} + 2 \cdot 2^{-3} + 2 \cdot 2^{-4} = 14/16$.

Case (c): Using similar reasoning, we can show $w(L_3) = 6 \cdot 2^{-3} = 6/8$.

Case (d): Likewise, we have $w(L_3) = 2 \cdot 2^{-2} + 3 \cdot 2^{-3} + 2^{-4} = 15/16$.

In all possible cases we have $w(L_3) < 1$. Thus, Ophelia can force a draw on Π_3. ∎

Draws on affine planes

Using the same technique, we can give the drawing strategy for Ophelia on infinitely many affine planes. For an affine plane of order n, L_1 consists of $n + 1$ partial lines of cardinality $n - 1$ (once X_1 is removed) and $(n^2 + n) - (n + 1)$ lines of cardinality n.

So, we have

$$w(L_1) = \sum_{i=1}^{n+1} 2^{-(n-1)} + \sum_{i=1}^{n^2-1} 2^{-n} = \frac{n^2+2n+1}{2^n}.$$

We see that $w(L_1) < 1$ when $n \geq 6$. Following the same argument as given in the previous proof, we see that Ophelia can force a draw on the affine planes of order $n \geq 7$ (since there is no such plane of order 6).

The only affine planes remaining are π_4 and π_5. It is interesting to note that we found greater difficulty determining the outcome of play on Π_3, π_4, and π_5 than on planes of higher order. While we were able to determine the outcome of play on Π_3 by performing calculations for all possible outcomes by hand, the unsuspected complexity of play on π_4 and π_5 lent itself to analysis by computer.

Ophelia's drawing strategy for π_5 is the same as that given for Π_3. The initial configurations are identical to the cases shown in FIGURE 8, and the weight functions for each case can be calculated as demonstrated in the previous proof. However, in π_5 some of these cases produced too many subcases to be calculated by hand, and a computer was used to verify that $w(L_i)$ was eventually less than 1. The following theorem summarizes these results.

DRAW THEOREM FOR π_n. *Ophelia can force a draw on every affine plane of order n with $n \geq 5$.*

There is only one plane left to consider. What happens on the affine plane of order 4? The following game shows that draws exist on π_4.

X	O	X	X
O	X	O	O
X	O	X	X
O	X	O	O

Since we had no examples of a plane for which a winning strategy and draws coexisted, it was natural to expect that Ophelia could force a draw. To our surprise, three independent computer algorithms show that Xeno has a winning strategy on this plane. The first two programs, written by students J. Yazinski and A. Insogna (University of Scranton), use a tree searching algorithm. The third program, written by I. Wanless (Oxford University), checks all possible games up to isomorphism. Thus, the affine plane of order 4 is the only plane for which Xeno has a winning strategy, and yet, draws exist. Finally, we have answered the two questions that we posed after initially introducing the game.

Answer 1: Xeno has a winning strategy on π_2, Π_2, π_3 and π_4.

Answer 2: Draws exist on π_n where $n \geq 4$, and on Π_n where $n \geq 3$.

Blocking configurations

Suppose you are playing as Ophelia on one of the infinitely many planes for which there is a drawing strategy. The algorithm given in the previous section may guarantee a draw, but it requires computations of Eulerian proportion in order to pick a point of maximum weight at each move. Since any opponent would surely cry foul were you to consult a computer, it could take *hours* to finish a game using this algorithm! The geometry of these planes suggests a more practical solution. We will relate Ophelia's strategy to this geometry in order to demonstrate some configurations (which Ophelia would like to construct) that can produce a draw with very few points. The desired set of points is called a blocking set, since every line intersects the set, but no line is contained in the set. More information on similar configurations can be found in recent survey articles [**6, 12**] and the references therein.

First, let us consider the projective plane of order 3. Since it is easier to understand Π_3 by describing P and L rather than giving its graph, take the elements of the following array on the left as the point set of this plane, and the right array as a possible game.

$$\begin{pmatrix} & & 1 & & \\ & 2 & 3 & 4 & \\ 5 & 6 & 7 & 8 & 9 \\ & 10 & 11 & 12 & \\ & & 13 & & \end{pmatrix} \qquad \begin{pmatrix} & & X & & \\ & X & X & X & \\ O & O & X & O & O \\ & X & O & X & \\ & & O & & \end{pmatrix}$$

We have simply taken the standard form of π_3 and added 1, 5, 9, and 13 as the points at infinity. The lines are given by

$$\{2,3,4,13\}, \{6,7,8,13\}, \{10,11,12,13\}, \{2,6,10,1\}, \{3,7,11,1\},$$
$$\{4,8,12,1\}, \{2,7,12,9\}, \{3,8,10,9\}, \{4,6,11,9\}, \{4,7,10,5\},$$
$$\{3,6,12,5\}, \{2,8,11,5\}, \quad \text{and} \quad \{1,5,9,13\}.$$

It is easily checked that the game shown on the right above is a draw. The set of points marked with X, $\{1, 2, 3, 4, 7, 10, 12\}$, and those marked with an O, $\{5, 6, 8, 9, 11, 13\}$, are both blocking sets. Further inspection shows that these sets have a specific configuration in common. We will focus on Ophelia's blocking set. The line $\ell = \{1, 5, 9, 13\}$ has all but one point labelled with an O. Through at least one of the points on ℓ marked with an O, say 5, there is a line $m = \{2, 5, 8, 11\}$ that also has all but one point marked with an O. We also see that lines ℓ and m contain five of the six points that compose Ophelia's blocking set. The sixth point lies on the line through points 1 and 2, the two points marked with an X on lines ℓ and m. We can find the same configuration in Xeno's blocking set by taking $\ell' = \{2, 3, 4, 13\}$ and $m' = \{2, 7, 12, 9\}$, which makes 5 the sixth point since it lies on the line through points 9 and 13. Of course, 10 is an extraneous point of the set for Xeno, included for the sake of presenting a complete game.

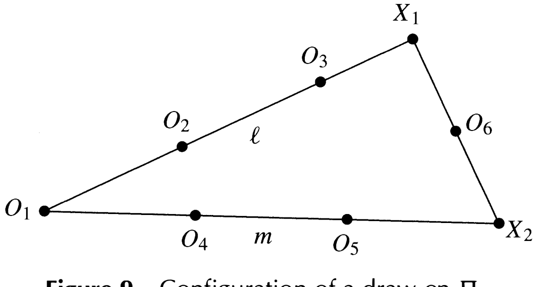


Figure 9 Configuration of a draw on Π_3

Interestingly, every draw on Π_3 displays such a configuration, as depicted in FIGURE 9. To show this, assume Ophelia has a produced a draw on Π_3 with the points in the set $A = \{O_1, \ldots, O_6\}$. If no three points of A are on a line, then through O_1 there is a line containing each O_i, $2 \leq i \leq 6$, and these five lines are distinct. However, this is impossible since there cannot be five lines through O_1 on Π_3. (Note also that no four points of A are on a line because then A would contain a line.) Hence, some three points of A are on a line. Without loss of generality, assume that we now have line ℓ as shown in FIGURE 9. X_1 has three lines through it other than ℓ. Each of these lines must have a point claimed by Ophelia since the set A has a point on every line. Hence, each of the remaining three points of A must be incident with exactly one of these lines, and a simple check shows that we have the configuration given above. This blocking configuration is not unique to Π_3. It can be generalized to projective planes of higher order as shown by the following theorem.

BLOCKING SETS ON Π_n THEOREM. *On any projective plane of order n with $n \geq 3$, there exists a blocking set of 2n points.*

Proof. This purely geometric result is shown within the game structure by constructing the blocking set. Let ℓ be a line in a projective plane of order $n \geq 3$, with points $q_1, \ldots, q_{n+1}$. Suppose Ophelia has accumulated $O_i = q_i$ for $i = 1, \ldots, n$. On Π_n, each of these points is incident with $n + 1$ lines. Hence $n^2 + 1$ lines now have an O on them.

Assume Xeno claims q_{n+1}, otherwise Ophelia wins. There are n lines through q_{n+1} other than ℓ. Label these lines $\ell_1, \ldots, \ell_n$, as in FIGURE 10. Choose a line $m \neq \ell$ through q_1 and let O_{n+i} be the intersection of m and ℓ_i for $i = 1, \ldots, n - 1$. Let O_{2n} be any point on ℓ_n other than the intersection of m and ℓ_n, otherwise Ophelia wins. Since $n > 2$, we are guaranteed that such a point exists. Finally, we have $\{O_1, O_2, \ldots, O_{2n}\}$ as the required set of $2n$ points since each of the $n^2 + n + 1$ lines are incident with a point in this set, and no line is contained in this set. ■

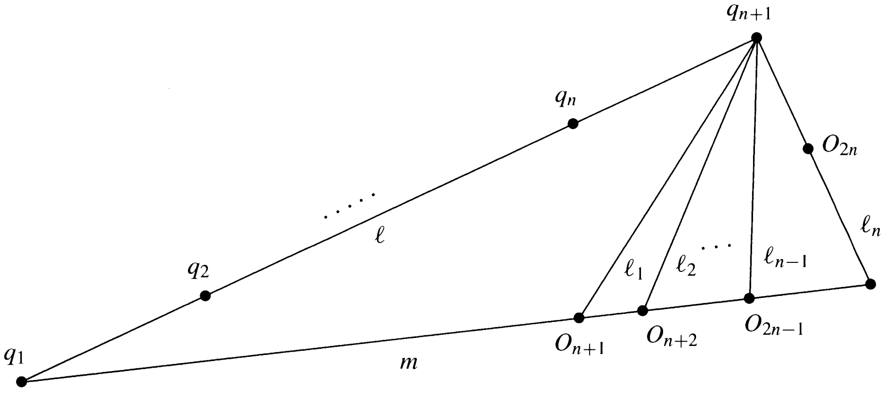


Figure 10 Configuration of blocking set on Π_n

The blocking set constructed in the proof translates to a drawing strategy for Ophelia that is free from computation. On a projective plane, Ophelia may attempt to acquire points that display such a configuration. At first consideration, the reader might find this strategy counterintuitive. If Ophelia's goal is to block every line, then how could it make sense to continue to place her marks on lines that are already blocked (ℓ and m)? The answer lies in the geometry of these planes. Since each point is incident with $n + 1$ lines, O_1 blocks $n + 1$ lines. Since there exists a line between O_1 and any other point, O_2 will block n lines regardless of its placement. O_3 will block n lines if it is placed on ℓ, but only $n - 1$ lines if not placed on ℓ. By continuing to place her marks on ℓ, Ophelia is maximizing the number of lines blocked by each O_i. The points $q_1, q_2, \ldots, q_n$ claimed by Ophelia, as shown in FIGURE 10, block $n^2 + 1$ of the $n^2 + n + 1$ lines on Π_n. This is the largest number of lines she can block with n points.

The blocking set on an affine plane of order greater than 4 displays a similar configuration, consisting of $2n - 1$ points. To show this, let $\ell_1, \ldots, \ell_n$ be the lines of a parallel class and suppose $\ell_1 = \{q_1, q_2, \ldots, q_n\}$, as in FIGURE 11. Let $O_i = q_i$ for $1, \ldots, n-1$, and assume Xeno claims q_n. Let $\ell_1, m_1, \ldots, m_n$ be the lines through q_n, and let O_{n-1+j} be the point of intersection of m_j and ℓ_{j+1} for $j = 1, \ldots, n-1$. Let O_{2n-1} be any point on m_n that is not collinear with $O_n, \ldots, O_{2n-2}$. This can be done because $n > 4$, that is, there are more than four points on a line.

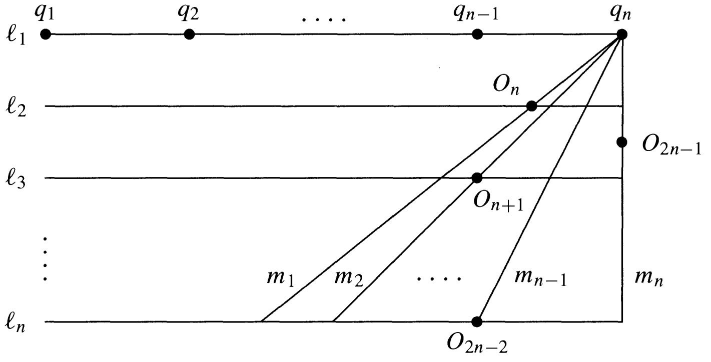


Figure 11 Configuration of blocking set on π_n

Notice that $O_n, O_{n+1}, \ldots, O_{2n-1}$ do not lie on a line since this line would have to interesect ℓ_1, which it does not. However it may be possible that $n - 1$ of them lie on a line with O_i for $i \leq n - 1$. This can be avoided by simply changing the order of lines $\{m_i\}$, which is possible when $n > 4$. Thus, we have $\{O_1, O_2, \ldots, O_{2n-1}\}$ as the required set of $2n - 1$ points. This work establishes the following result.

BLOCKING SETS ON π_n THEOREM. *On any affine plane of order n with $n \geq 5$, there exists a blocking set of* $2n - 1$ *points.*

As we can see from these proofs, if Ophelia can place $2n$ or $2n - 1$ marks (depending on the type of plane) in the required manner then the game will be a draw. While this offers the second player an easy algorithm to follow, it is not a drawing strategy since it is not guaranteed to produce a draw. If Xeno's best move happens to be a point on the line which was to be part of Ophelia's blocking configuration, then she must begin acquiring points on a different line.

Competitive play

At the University of Scranton we hold an annual single-elimination "Tic-tac-toe on π_4" tournament where students compete for the top spot (and prizes!). It is not uncommon to see the serious competitors practicing for weeks before the contest. We encourage the reader to play too, as we have found that these students gain not only an understanding of affine planes, but also develop an intuition for finite geometries that reveals properties and symmetries not easily seen by reading definitions in a geometry text. For practice on π_4, try playing against a computer at the author's website: `academic.uofs.edu/faculty/carrollm1/tictactoe/tictactoea4.html`. We also welcome a proof of the existence of a winning strategy for Xeno on π_4 that does not rely on a computer. For further reading, surveys of other tic-tac-toe games can be found in Beck [**2**] and Berlekamp, Conway, and Guy [**4**].

Acknowledgments. We thank Jonathan Yazinski, Al Insogna, and Ian Wanless for their enthusiasm in determining the outcome of play on π_4.

REFERENCES

1. E. F. Assmus Jr. and J. D. Key, *Designs and their Codes*, Cambridge University Press, Cambridge, UK, 1992.
2. J. Beck, Achievement Games and the Probabilistic Method, *Combinatorics, Paul Erdős is Eighty*, Vol. **1**, Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 1993, 51–78.
3. M. K. Bennett, *Affine and Projective Geometry*, Wiley, New York, 1995.
4. E. R. Berlekamp, J. H. Conway, and R. K. Guy, *Winning Ways for Your Mathematical Plays*, Vol. **2**, Academic Press Inc., London-New York, 1982.
5. K. Binmore, *Fun and Games: A Text on Game Theory*, D. C. Heath and Co., Lexington, MA, 1992.
6. A. Blokhuis, Blocking sets in Desarguesian planes, *Combinatorics, Paul Erdős is Eighty*, Vol. **2**, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996, 133–155.
7. R. C. Bose, On the application of the properties of Galois fields to the problem of construction of hyper-Graeco-Latin squares, *Sankhya* **3** (1938), 323–338.
8. S. T. Dougherty, A coding theoretic solution to the 36 officer problem, *Des. Codes Cryptogr.* **4** (1994), 123–128.
9. P. Erdős and J. L. Selfridge, On a combinatorial game, *J. Combin. Theory* **14** (1973), 298–301.
10. L. Euler, Recherches Sur une nouvelle espace de quarees magiques, *Verh. Zeeuwsch Genootsch. Wetensch. Vlissengen* **9** (1782), 85–239. Reprinted in L. Euler, *Opera Omnia*, ser. 1, vol. 7, Tuebner, Berlin-Leipzig, 1923, 291–392.
11. A. W. Hales and R. I. Jewett, Regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963), 222–229.
12. J. W. P. Hirschfeld and L. Storme, The packing problem in statistics, coding theory, and finite projective spaces, R. C. Bose Memorial Conference (Fort Collins, CO, 1995), *J. Statist. Plann. Inference* **72**:1–2 (1998), 355–380.
13. C. Lam, The search for a finite projective plane of order 10, *Amer. Math. Monthly* **98** (1991), 305–318.
14. C. F. Laywine and G. L. Mullen, *Discrete Mathematics Using Latin Squares*, J. Wiley and Sons, New York, 1998.
15. X. Lu, A characterization on n-critical economical generalized tic-tac-toe games, *Discrete Math.* **110** (1992), 197–203.
16. A. Rapoport, *Two-Person Game Theory; The Essential Ideas*, University of Michigan Press, Ann Arbor, 1966.
17. F. W. Stevenson, *Projective Planes*, W. H. Freeman and Co., San Francisco, 1972.
18. D. R. Stinson, A short proof of the nonexistence of a pair of orthogonal Latin squares of order six, *J. Combin. Theory Ser. A* **36** (1984), 373–376.
19. G. Tarry, Le problème des 36 officers, *Compte Rendu Ass. Franc. Pour l'avancement des Sciences* **2** (1901), 170–203.
20. J. Weeks, *Torus and Klein Bottle Games: Tic-Tac-Toe*, available at `http://www.geometrygames.org/TorusGames/html/TicTacToe.html`.

Cover image: *Where did you say that camera was?*, by Don MacCubbin

Jim Henle's article in this issue tells us how to compute the location from which a given photograph was taken. This is a potentially confusing problem, but the mathematician shown on the cover will solve it quickly after reading Henle's article.

Don MacCubbin is an artist, a photographer, and the Mechanical Engineering Lab Manager at Santa Clara University, where he just received his bachelors' degree in Studio Art. When he isn't busy with undergrads in the lab, Don ponders triangulation as it applies to errant golf balls and missed pool shots.

Designs, Geometry, and a Golfer's Dilemma

KEITH E. MELLINGER
Mary Washington College
Fredericksburg, VA 22401
kmelling@mwc.edu

I was taken off guard the other day when my father-in-law, John, posed to me a very simply stated problem. He plays golf. In fact, John plays a lot of golf. When you play as much golf as he does, you become bored playing with the same people over and over again. So here's the problem: John regularly plays with a group of 16 people. Three days a week for the entire summer, they go out in 4 groups of 4 players each to hit the course. Is there some way they can arrange the players in the groups each day so that everybody plays with everybody else in some sort of regular way? As my father-in-law said, "We want to mix it up as much as possible."

First of all, we need to figure out what the question is asking. Let's look at the problem from the perspective of my father-in-law. Suppose that John plays his first day with three other players, say Keith, Bill, and Howard. Then, there are still 12 other people available to play. John would prefer to play with all 12 other people before he ends up playing with Keith, Bill, or Howard again. From John's perspective, this seems like it may not be a difficult problem. Simply assign three players to John for the first day, three different players to John for the second day, etc. Then, after 5 days, John will have played with all of the other 15 people in the group. However, remember that we need to assign 4 groups (not just John's group) of 4 players each, and we want *every* golfer to play with *every other* golfer, again in some sort of regular way. Suddenly the problem seems much more difficult. In the next several pages, we will find some solutions to the problem. In our quest to find a best solution we will take a ride through some areas of discrete mathematics including finite affine and projective planes, and combinatorial designs.

A connection to affine planes

One basic solution to the golfer's dilemma comes from a rather unexpected area of mathematics, geometry. How could this be relevant? Bear with me for a few paragraphs and we'll get back to golfing soon enough. We need some terminology. The formal definition of an affine plane goes like this.

DEFINITION. An **affine plane** is a set of points together with a collection of subsets of these points, called lines, such that

1. every two distinct points determine a unique line,
2. if l is a line and P is a point not on l, then there exists a unique line m such that P is on m and l and m have no points in common, and
3. there exist 3 noncollinear points.

(You can read about playing tic-tac-toe on affine planes in Carroll and Dougherty's article in this issue of the MAGAZINE.) One important point, which is made clear by the second axiom, is that affine planes have *parallel lines*. This may not seem like a big deal, but in the world of higher mathematics, people do without parallel lines all the time. In fact, we will soon see another kind of plane where parallel lines do not exist. Now we add an additional condition. Suppose that we have an affine plane $\mathcal{A}$ that contains only a finite number of points. Is this possible? Indeed it is.

Consider a 2-dimensional vector space V over some field $\mathcal{F}$. We define points to be all of the vectors of V and define lines to be all of the cosets of all of the 1-dimensional subspaces contained in V. For example, take V to be the vector space $\mathbb{R}^2$ whose vectors are all ordered pairs (x, y) for $x, y \in \mathbb{R}$. The cosets of the 1-dimensional subspaces are sets of the form $\{\mathbf{u} + t\mathbf{v} : t \in \mathbb{R}\}$ for some $\mathbf{u}$, $\mathbf{v}$ in V. These cosets of V are exactly what we typically call the lines of the coordinate plane. Hence, a 2-dimensional vector space can be used to model an affine plane.

Now, we again consider V as a 2-dimensional vector space, but this time restrict the coordinates of the vectors of V to be in the finite field $GF(q)$ that contains q elements. The notation $GF(q)$ means the *Galois field* with q elements, named after the French mathematician Evariste Galois (1811–1832). For those unfamiliar with finite fields, simply think of the coordinates x and y as coming from a finite set that only contains q elements. One can prove that the number of elements in a finite field is always a power of some prime number. Hence, we refer to q as being a *prime power*. Again, considering all of the vectors of V as points and all of the cosets of all of the 1-dimensional subspaces as lines, we obtain an affine plane. This time, however, our affine plane contains only a finite number of points, namely, the number of vectors of V. Since each vector is written as an ordered pair of elements from $GF(q)$, we see that V contains q^2 vectors. By varying t in the definition of cosets given above, we see that the number of points on a line is equal to the number of elements in the finite field. Hence, every line contains exactly q points.

We can do some more involved counting to find other properties of our affine plane. For instance, fix a vector $\mathbf{v} \in V$ and count how many cosets of 1-dimensional subspaces pass through $\mathbf{v}$. To do this, we note that there are $q^2 - 1$ choices for a second vector $\mathbf{w}$ different from $\mathbf{v}$. The vectors $\mathbf{v}$ and $\mathbf{w}$ together determine a coset of a 1-dimensional vector subspace, say C. But this coset C could be determined from $\mathbf{v}$ and *any* other vector in C. Since there are $q - 1$ choices for another vector in C, each such coset has been counted $q - 1$ times. Therefore, the total number of cosets of 1-dimensional subspaces through the given vector $\mathbf{v}$ is exactly $(q^2 - 1)/(q - 1) = q + 1$. Hence, every point lies on exactly $q + 1$ lines.

Finally, the number of lines can be counted by counting the number of ways to choose two distinct points to generate a line, and then dividing by the number of ways any given line was counted. The number of ways to choose an ordered pair of two distinct vectors is $q^2(q^2 - 1)$. But each line is generated by choosing *any* such pair of points on that line, which can be done in $q(q - 1)$ ways. Hence, the total number of lines is exactly $(q^2(q^2 - 1))/(q(q - 1)) = q^2 + q$. The affine plane obtained from this model of a 2-dimensional vector space over the finite field $GF(q)$ is denoted $AG(2, q)$ (the classical affine geometry of dimension 2 and order q).

We can say a little more. Note that two cosets of the same 1-dimensional subspace of a vector space never intersect. Hence, we have collections of lines in our affine plane no two of which meet. Such sets of lines are naturally called *parallel classes of lines*. Counting can again be used to show that each parallel class contains exactly q lines. Since there are a total of $q(q + 1)$ lines, there must be $q + 1$ different parallel classes.

Let's get back to the original problem of the golfer's dilemma. We have a total of 16 golfers that we want to break into various groups of 4. Now, let $q = 4$ in the affine plane model above. The affine plane $AG(2, 4)$ contains exactly 16 points, and every line contains exactly 4 points. Every parallel class contains exactly 4 lines, and there are exactly 5 parallel classes. Hence, we have a solution to the golfer's dilemma by letting the points of $AG(2, 4)$ represent the golfers, and the lines represent the various groups of 4 golfers playing together. The parallel classes of lines represent the various days of play since each parallel class consists of 4 distinct groups of 4 players each (that is, 4 parallel lines in each parallel class).

When the finite field is relatively small, one can try to find the lines of $AG(2, q)$ by hand. A software package such as *Magma* [**3**] does this computation virtually instantaneously, but since $q = 4$ is pretty small, let's get started doing it by hand. Keep in mind that since this is a *finite* affine plane, lines can be thought of simply as subsets of points with little relation to shapes. First we write the 16 players in a 4 × 4 grid as in FIGURE 1.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Figure 1 Representing $AG(2, 4)$

The rows and columns of the grid can each represent a parallel class. This means that any further lines must contain exactly one point from each row, and one point from each column (since two points in the same row or column would uniquely determine one of the lines already given). At this point we might try using the diagonals to get two more lines. The reader is encouraged to try to find the remaining lines by hand before proceeding, the lesson being that *this is not at all easy*. The use of finite geometry (along with a little computer power to generate the cosets of the appropriate vector subspace) gives us the blocks in TABLE 1.

TABLE 1: A five-day schedule for 16 golfers.

Day 1	Day 2	Day 3	Day 4	Day 5
(rows)	(columns)	(diagonals)		
$\{1, 2, 3, 4\}$	$\{1, 5, 9, 13\}$	$\{1, 6, 11, 16\}$	$\{1, 7, 12, 14\}$	$\{1, 8, 10, 15\}$
$\{5, 6, 7, 8\}$	$\{2, 6, 10, 14\}$	$\{2, 5, 12, 15\}$	$\{2, 8, 11, 13\}$	$\{2, 7, 9, 16\}$
$\{9, 10, 11, 12\}$	$\{3, 7, 11, 15\}$	$\{3, 8, 9, 14\}$	$\{3, 5, 10, 16\}$	$\{3, 6, 12, 13\}$
$\{13, 14, 15, 16\}$	$\{4, 8, 12, 16\}$	$\{4, 7, 10, 13\}$	$\{4, 6, 9, 15\}$	$\{4, 5, 11, 14\}$

We have solved the problem of the golfer's dilemma: Assign each golfer a number between 1 and 16. Then, over the course of 5 days, the golfers play together based on the schedule outlined in TABLE 1. At the end of 5 days, every golfer will have played with every other golfer exactly once.

The golfers aren't happy

Having found a solution to the problem, I was quick to email a solution to my father-in-law, but was rather disappointed at his immediate response. First of all, we have only covered 5 days of play. These guys want to play *all summer*. So what do we do? A natural remedy is to simply repeat the process. That is, after 5 days, just start over with day 1. That way, after, say, 25 days of play, every golfer will have played with every other golfer exactly 5 times. However, there is a clear disadvantage to repeating our solution.

Let's go back to the beginning. Suppose that John plays his first day with Keith, Bill, and Howard. Then, 5 days later, the golfers all decide to repeat the schedule. When

John plays with Keith for the second time, the other two members of their group will again be Bill and Howard. It would be nice if John and Keith could play together with two *different* people the next time around. More precisely, we see that two distinct golfers uniquely determine a group. That is, if I pick any two golfers, say John and Keith, from the group of 16, there is exactly one group of 4 in which John and Keith are both members. In our example, it is the group that contains Bill and Howard. From the perspective of the affine plane, this really comes as no surprise. Recall that two points of the affine plane determine *exactly one* line.

So let's kick it up a notch. Here's one quick and easy way to remedy the situation. Assign each golfer a number between 1 and 16, and play through the five day schedule as outlined in TABLE 1. After the five days are up, permute the numbers in some way, and then repeat the schedule. The golfers could all pick a partner to switch numbers with, or they could cyclically shift their numbers ($i \rightarrow i+1$ for i between 1 and 15, and $16 \rightarrow 1$). Of course, one must be careful with such a cyclic shift. The reader should check that after certain cyclic shifts, the groups will start to repeat. Is there some more systematic way to ensure that every golfer plays with every other golfer, but eliminate the drawbacks of the solution already given?

Statisticians face these sorts of questions all the time when they are designing experiments. They have a set of v objects on which they want to run an experiment, but the experiment can only be run on k objects at a time. In our case, $v = 16$ and $k = 4$, and maybe our experiment consists of determining the ability of each golfer. The statisticians want to mix things up as much as possible. Maybe object 2 could affect the outcome of the experiment on object 1; maybe Bill makes John nervous. So, in order to get an accurate reading on John's golf ability, we need to make sure that Bill doesn't play with John every single time. In a similar fashion, suppose Bill alone doesn't make John nervous, but when Bill and Howard get together, they goof around a lot and it makes John nervous. So, it would be OK to put these three together once, but if John and Bill are together again, it would be best if Howard isn't included the second time around. More generally, *we would like every set of three golfers to be grouped together exactly once*. Can this be done?

First we note that if every two golfers are together exactly once, then the schedule would run in 5 days (as discussed above). This makes sense simply by counting. That is, if John plays with 3 different people each day, it would take 5 days for him to play with all of the remaining 15 players. Can we apply the same reasoning to the new problem? That is, suppose every group of 3 golfers play together exactly once. How long will the schedule last? From John's perspective, the answer is equal to the number of groups in which John is a member. But remember, *three* golfers determine a group now. The number of ways to choose 2 golfers from the remaining 15 is $\binom{15}{2} = 105$. Once two other golfers are chosen, John and the two others uniquely determine the group, say, John, Keith, Bill, and Howard. But whether we pick Keith and Bill, Keith and Howard, or Howard and Bill as the additional two golfers, we will always get the same group. Hence, each such group is counted 3 times. Therefore, the number of groups in which John is a member is $105/3 = 35$. So the schedule would last for 35 days, or about 12 weeks if they play 3 days per week. This would cover most of the summer and probably keep the golfers (in particular, my father-in-law) happy.

Combinatorial designs

Mathematicians refer to the solution of a problem similar to the one above as a *combinatorial design*, or simply a *design*.

DEFINITION. A **design** is a set of v *points* together with a set of subsets of size k of these points, called *blocks*, with the property that any t points lie in exactly λ blocks. Such a design with these parameters is called a $t - (v, k, \lambda)$ design.

That's a lot of variables. Let's look at an example. In our first solution to the problem, we had 16 golfers playing in groups of 4 such that every pair of golfers played together exactly once. Hence, the number of golfers $= v = 16$, the size of the groups (or blocks) $= k = 4$, and every $t = 2$ golfers play together exactly $\lambda = 1$ times. The affine plane model provided us with a $2 - (16, 4, 1)$ design that solved the problem.

Based on our discussion in the last section, we now desire a $3 - (16, 4, 1)$ design. That is, we want every three golfers to play together exactly once. Further, we would like to assign 4 pairwise disjoint groups to each day for 35 days. So, not only do we need to build a $3 - (16, 4, 1)$ design, but we need to be able to divide the blocks of the design into 35 sets of 4 pairwise disjoint blocks each (a design with this property is called *resolvable*). It sounds like a big task, but it turns out that the design we seek was actually discovered many decades ago. One excellent source for such information is the *CRC Handbook of Combinatorial Designs* [**4**]. It is here that you can find all known values of t, v, k, and λ for which a design exists. The existence of the design we seek is due to Hanani [**5**]. However, the construction of this design relies on first finding a $3 - (8, 4, 1)$ design (that is, finding an equivalent golf schedule for only 8 golfers rather than 16). Oddly enough, the construction of this smaller design also has a connection to geometry.

Projective planes

There is a close connection between affine planes and the so-called *projective planes*. Projective planes correspond to the notion of perspective. That is, from the perspective of a man standing on railroad track, the tracks seems to meet out at the horizon. Hence, parallel lines do not seem to exist. This can be laid out mathematically as follows.

DEFINITION. A **projective plane** is a set of points, together with a set of subsets of these points, called lines, such that

1. every two distinct points determine a unique line,
2. every two distinct lines meet in a unique point, and
3. there exist four points, no three of which are collinear.

Just as we did with the affine plane, we can use a vector space to model a projective plane. This time, we start with a 3-dimensional vector space V over some field $\mathcal{F}$. We take as our points the 1-dimensional subspaces of V. The 2-dimensional subspaces of V are our lines. Since two distinct 1-dimensional subspaces determine a unique 2-dimensional subspace, axiom 1 is satisfied. Similarly, two distinct 2-dimensional subspaces meet in a unique 1-dimensional subspace. Hence, axiom 2 follows. Finally, we can easily find vectors to satisfy axiom 3. For instance, we could use the vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, and $(1, 1, 1)$.

For our purposes, we only need one specific projective plane. Referring to the vector space model above, it would correspond to a 3-dimensional vector space over the finite field with only 2 elements, $GF(2)$. This is probably the most famous projective plane and is more commonly known as the Fano plane. It contains 7 points, 7 lines (one of which is represented by the circle), and every line contains exactly 3 points (see FIGURE 2).

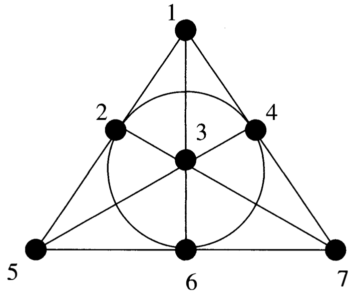


Figure 2 The Fano plane

From the Fano plane, we get a $2-(7,3,1)$ design by letting the lines of the plane represent the blocks of our design (see [2] for much more on this famous design). That's not quite what we want. Recall that we are looking for a design on 8 points, not 7, in order to eventually build the larger design on 16 points. We can use the Fano plane to build the design we need.

Label the points of the Fano plane with the integers 1 through 7 and consider the set of these points together with one additional point labelled 8. These will be the points of our new design. The blocks for our new design are of two types. The first type of block is a line of the Fano plane together with the extra point 8. The second type of block is any set of four points of the Fano plane such that no three of the points are collinear. Such a set of points is known as a *hyperoval.* For instance, in the labelling in FIGURE 2, note that points 1, 2, 3, and 4 form a set of 4 points, no 3 of which lie on a common line. Hence, these points form a hyperoval. Enumerating all such hyperovals and combining these with the other type of blocks defined above, we obtain the 14 blocks in TABLE 2.

TABLE 2: Blocks of the $3-(8,4,1)$ design.

1	$\{1,2,5,8\}$	$\{3,4,6,7\}$
2	$\{1,3,6,8\}$	$\{2,4,5,7\}$
3	$\{1,4,7,8\}$	$\{2,3,5,6\}$
4	$\{2,3,7,8\}$	$\{1,4,5,6\}$
5	$\{2,4,6,8\}$	$\{1,3,5,7\}$
6	$\{3,4,5,8\}$	$\{1,2,6,7\}$
7	$\{5,6,7,8\}$	$\{1,2,3,4\}$

Note that we can write the blocks in a table so that any two blocks in a row are disjoint. One can easily check that any three of our points (the points of the Fano plane, plus the additional point 8) lie together in exactly one block from TABLE 2. Hence, we have constructed a $3-(8,4,1)$ design. Moreover, we have solved the golfer's dilemma in the case when there are 8 golfers. That is, we have constructed a 7-day schedule (the rows of TABLE 2) in which every 3 golfers will play together exactly once.

A better solution

We can use the $3-(8, 4, 1)$ design to build the $3-(16, 4, 1)$ design we seek. For each row of TABLE 2, we will construct a golf schedule for 5 days, thereby giving us the 35 day schedule we need. Let $B = \{a, b, c, d\}$ be any block from TABLE 2. Then the block B constructs 10 new blocks in the manner shown in TABLE 3.

TABLE 3: 10 new blocks from the old block $\{a, b, c, d\}$.

1	$\{a, b, c, d\}$	$\{a+8, b+8, c+8, d+8\}$
2	$\{a+8, b+8, c, d\}$	$\{a, b, c+8, d+8\}$
3	$\{a+8, b, c+8, d\}$	$\{a, b+8, c, d+8\}$
4	$\{a+8, b, c, d+8\}$	$\{a, b+8, c+8, d\}$
5	$\{a, b, a+8, b+8\}$	$\{c, d, c+8, d+8\}$

So each block of the old design from the Fano plane is used to construct 10 new blocks of the design we seek. Hence, we obtain $14 \cdot 10 = 140$ new blocks. All we need now is to partition these 140 blocks into 35 sets (representing the days) of 4 blocks each (representing the groups of golfers).

We are finally ready to construct our solution to the golfer's dilemma. We construct the 4 sets of 4 golfers each for any particular day by first choosing a row from TABLE 2, and then constructing the four associated groups using a row from TABLE 3. For instance, if we select row 4 of TABLE 2 and row 2 of TABLE 3, we obtain the four groups:

$$\{10, 11, 7, 8\}, \{9, 12, 5, 6\}, \{2, 3, 15, 16\}, \{1, 4, 13, 14\}.$$

It is not too difficult to see that this construction gives us what we want. First note that any particular day partitions the 16 golfers into four groups of four since the groups in any row of TABLES 2 and 3 are disjoint. In addition, 3 points of the new design determine a unique block since 3 points from the $3-(8, 4, 1)$ design determine a unique block. This follows since we always alter an even number of entries in the original blocks to obtain the new blocks. As a result, for any given triple $\{a, b, c\}$, we can always backtrack through the tables to find the block that contains a, b, and c. This shows that we indeed have a $3-(16, 4, 1)$ design.

For instance, suppose we want to find the unique block containing $\{1, 11, 14\}$. First, we reduce the integers by subtracting 8 from any value larger than 8 and label the results as $a = 1$, $b = 3$, and $c = 6$. Next, we look for the row in TABLE 2 containing $\{1, 3, 6\}$ as a subset of a block. This is row 2. Now we look for the row in TABLE 3 that will keep a unaltered, but adds 8 to b and c. This is row 4. Hence, the day corresponding to rows 2 and 4 of the two tables (respectively) has golfers 1, 11, and 14 playing together (with golfer 8).

Note that taking all possible combinations of rows of the two tables gives us the 35 day schedule we desire. Hence, we can construct a 35 day schedule for the 16 golfers such that every group of three golfers will play together in a group exactly once.

Can we go further?

Is this the best we can do? Let's think about extending our old argument. Recall that John got nervous when he played with both Bill and Howard and that three golfers uniquely determine a group. Hence, when John, Bill, and Keith play together, the fourth golfer, say Howard, is uniquely determined. Suppose that John, Bill, and Keith enjoy playing together, but do not necessarily want Howard as their fourth every time they are together. What are we saying? Essentially, *we want every possible combination of 4 players to be together exactly once*. Is it unrealistic to ask for such an extreme condition?

Let's start with some simple counting. Again, we look at everything from John's perspective. If he plays with every possible combination of 3 other golfers, then he would have to play exactly $\binom{15}{3} = 455$ times. This would certainly not be obtainable in a summer! But mathematically, it certainly seems possible and it is (see Theorem 38.1 in [**6**]). Realistically, a solution that takes this much time to complete would probably not be feasible for the average golfer. Hence, in my opinion, the solution given in the previous section is the best possible. My father-in-law seemed to like it too.

For more on projective and affine geometry and its connections to some modern problems in design theory, as well as the theory of error correcting codes and cryptography, you may want to check out *Projective Geometry* by Beutelspacher and Rosenbaum [**1**].

REFERENCES

1. A. Beutelspacher and U. Rosenbaum, *Projective Geometry: From Foundations to Applications*, Cambridge University Press, Cambridge, UK, 1998.
2. E. Brown, The many names of (7, 3, 1), this MAGAZINE **75** (2002), 83–94.
3. J. Cannon and C. Playoust, *An Introduction to Magma*, University of Sydney, Sydney, Australia, 1994.
4. *CRC Handbook of Combinatorial Designs*, ed. C. J. Colbourn and J. H. Dinitz, CRC Press, Boca Raton FL, 1996.
5. H. Hanani, On quadruple systems, *Canad. J. Math.* **12** (1960), 145–157.
6. J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, 2nd ed., Cambridge University Press, Cambridge, UK, 2001.

50 Years Ago in the MAGAZINE

From the preface of *Theory of Functions of a Complex Variable*, Vol. 1, by Constantin Caratheodory, New York, Chelsea Publishing Company, 1954, 314 pp., $4.95, quoted as part of a posthumous review of the book in Vol. **28**, No. 2, (Nov.–Dec., 1954), 122:

> The book begins with a treatment of Inversion Geometry (geometry of circles). This subject, of such great importance for Function Theory, is taught in great detail in France, whereas in German-language and English-language universities it is usually dealt with in much too cursory a fashion. It seems to me, however, that this branch of geometry forms the best avenue of approach to the Theory of Functions; it was, after all, his knowledge of Inversion Geometry that enabled H. A. Schwarz to achieve all of his celebrated successes.

Arithmetic Progressions with Three Parts in Prescribed Ratio and a Challenge of Fermat

KENNETH FOGARTY
CORMAC O'SULLIVAN
Bronx Community College
City University of New York
Bronx, NY 10453
kwfogarty@dmcom.net
cormac.osullivan@bcc.cuny.edu

The arithmetic progression 1, 2, 3 can be broken into two consecutive pieces that have equal sums by the relation $1 + 2 = 3$. The first author, in the problem pages of journals [**13, 14**], wondered if an arithmetic progression could be found that breaks into *three* consecutive pieces with equal sums. Here are some examples that come close:

$$4+5+6 = 7+8 = (9+10+11)/2,$$

$$3+5+7+9 = 11+13 = (15+17+19+21)/3,$$

$$(6+7+8+9)/2 = (10+11+12+13+14)/4 = 15.$$

This appealing question has a simple answer that turns out to be related to a certain Diophantine equation considered by Euler, namely

$$x^4 - x^2y^2 + y^4 = z^2 \tag{1}$$

where we are looking for integer solutions. In turn, (1) is related to the possibility of finding four squares as the consecutive terms of an arithmetic progression, a challenge issued by Fermat in 1640. We'll follow this thread and further address the question of arithmetic progressions with three parts in other fixed ratios. We close the article with four open questions, which we hope the reader will take as an invitation to further explore some of the mysteries of Diophantine equations.

Reduction to a Diophantine equation

So far we have been talking about sequences of integers. We may just as easily ask these questions for arithmetic progressions of real numbers. By an n-term arithmetic progression we therefore mean real numbers $e_1, e_2, \dots, e_n$ with common difference $e_{i+1} - e_i = \Delta > 0$ for $1 \le i < n$. If $n = a + b + c$, with positive integers a, b, c, we give names to the sums of the first a, the middle b, and the final c terms:

$$S_1 = \sum_{i=1}^{a} e_i, \quad S_2 = \sum_{i=a+1}^{a+b} e_i, \quad S_3 = \sum_{i=a+b+1}^{n} e_i. \tag{2}$$

The question we address in this article is: What are the possibilities for the ratios $S_1 : S_2 : S_3$? In particular, as we investigate in this section, can we ever have $S_1 = S_2 = S_3$? Clearly dividing each term in an arithmetic progression by the same number does not alter the ratios $S_1 : S_2 : S_3$ so after dividing by Δ we may make the simplifying assumption that the common difference of our progressions is always 1.

Using $1+2+\cdots+n = n(n+1)/2$ we have

$$2S_1 = a(2e_1 - 1 + a) = a(2e_{a+1} - 1 - a),$$
$$2S_2 = b(2e_{a+1} - 1 + b) = b(2e_{a+b+1} - 1 - b),$$
$$2S_3 = c(2e_{a+b+1} - 1 + c).$$

Setting $S_1 = S_2$ we find

$$2e_{a+1} - 1 = \frac{a^2+b^2}{a-b}. \tag{3}$$

Similarly $S_2 = S_3$ implies

$$2e_{a+b+1} - 1 = \frac{b^2+c^2}{b-c}. \tag{4}$$

Since $e_{a+b+1} = e_{a+1} + b$ we may solve for e_{a+1} in equations (3) and (4) to get

$$2b = \frac{b^2+c^2}{b-c} - \frac{a^2+b^2}{a-b}.$$

Rearranging we obtain the relation

$$ab^2 + a^2b + bc^2 + b^2c - ac^2 - a^2c - 2abc = 0. \tag{5}$$

Note that if any two of the positive integers a, b, c are equal then (5) implies that all three must be equal. Therefore by (3) and (4) we must have a, b, c all distinct.

PROPOSITION 1. *There exists an arithmetic progression with beginning, middle, and end having equal sums (with a, b, and c terms respectively) if and only if there exist positive distinct integers a, b, c satisfying equation* (5).

Proof. We have proved one direction. In the other, given such a, b, c let e_{a+1} be the rational number satisfying equation (3) and set $e_1 = e_{a+1} - a$. Then, as we have seen, the arithmetic progression $e_1, e_1+1, \ldots, e_1+a+b+c-1$ has the desired property. Also note that, if we like, we can make each term an integer by multiplying by $2(a-b)$. This completes the proof. ■

Let us therefore try to find integers a, b, c satisfying (5). Solving for b, we get

$$b^2(a+c) + b(a-c)^2 - ac(a+c) = 0, \tag{6}$$

a quadratic equation with discriminant

$$\delta = (a-c)^4 + 4ac(a+c)^2 = a^4 + 14a^2c^2 + c^4.$$

Set $p = a+c$ and $q = a-c$; then $\delta = q^4 + p^2(p^2-q^2)$. We have

$$b = \frac{\sqrt{\delta} - q^2}{2p},$$

which implies that we must have

$$p^4 - p^2q^2 + q^4 = r^2 \tag{7}$$

for some r. Conversely if p and q are integers satisfying (7) then

$$a = p(p+q), \quad b = \sqrt{p^4 - p^2q^2 + q^4} - q^2, \quad c = p(p-q)$$

are easily shown to satisfy (6). In this way (5) and (6) have solutions if and only if (7) does.

Equations like this one, where we seek only integer or only rational solutions, are called Diophantine equations in honor of Diophantus of Alexandria. Diophantus, who is thought to have lived in the third century [**1**], wrote the *Arithmetica*, where many such equations are solved.

In fact, $p^4 - p^2q^2 + q^4$ can be a square only if $p = \pm q$ or $pq = 0$. This is a result of Euler [**7**] from the 18th century. For completeness, we include an elegant proof of this fact by "infinite descent," which is due to Pocklington [**15**]. The result is also mentioned in Dickson's encyclopedic *History of the Theory of Numbers* [**4**, p. 638].

This method of proof, first employed by Fermat, is very useful in proving *negative* statements, for instance, that a certain equation has no (or only trivial) integer solutions. As we shall see, from an assumed initial solution to an equation, a new, strictly smaller, solution is constructed. Repeat the argument and an infinite chain of solutions, descending in size, appears. But this contradicts the fact that our solutions are bounded positive integers and hence finite in number. Thus our initial assumption of a solution to the equation was false.

Pocklington's proof uses the following well-known parameterization of Pythagorean triples: Let $x^2 + y^2 = z^2$ for positive integers x, y, z with $\gcd(x, y) = 1$. Necessarily one of x, y (say y) is even and there exist integers u, v with $\gcd(u, v) = 1$, $u > v > 0$ such that

$$x = u^2 - v^2, \quad y = 2uv, \quad z = u^2 + v^2.$$

For a proof see the classic text by Hardy and Wright [**9**, Theorem 225].

PROPOSITION 2. *If $p^4 - p^2q^2 + q^4 = r^2$ for positive integers p, q, r, then $p = q$.*

Proof. Assume that $p, q > 0$ are integer solutions to the above equation with $\gcd(p, q) = 1$. Suppose also that q is even (we will treat the case of p, q odd later) and that pq is minimal among all integer solutions. We have

$$(p^2 - q^2)^2 + (pq)^2 = r^2 \tag{8}$$

and $\gcd(p^2 - q^2, pq) = 1$ so that

$$p^2 - q^2 = u^2 - v^2, \tag{9}$$

$$pq = 2uv. \tag{10}$$

Considering the first equation (9) modulo 4 we see that v is even. In plainer language, since a square must have remainder 0 or 1 when divided by 4, the only possibility is that v^2 is divisible by 4. Next let

$$\alpha = \gcd(p, u), \quad \beta = \gcd(p, v), \quad \gamma = \gcd(q, u), \quad \delta = \gcd(q, v)$$

with α, β, γ odd and δ even. We have by (10)

$$p = \alpha\beta, \quad q = 2\gamma\delta, \quad u = \alpha\gamma, \quad v = \beta\delta.$$

Putting these back into (9) we obtain

$$\beta^2(\alpha^2 + \delta^2) = \gamma^2(\alpha^2 + 4\delta^2). \tag{11}$$

We want to demonstrate next that $\gcd(\alpha^2 + \delta^2, \alpha^2 + 4\delta^2)$ equals 1 or 3. To see this suppose d divides both $A = \alpha^2 + \delta^2$ and $B = \alpha^2 + 4\delta^2$. Then d will be a divisor of $B - A = 3\delta^2$ and $4A - B = 3\alpha^2$. Since α and δ are relatively prime d must be a factor of 3. Taking d as large as possible shows that $\gcd(\alpha^2 + \delta^2, \alpha^2 + 4\delta^2)$ is a factor of 3 as we said. But it cannot be 3 since 3 does not divide $\alpha^2 + \delta^2$ (squares must have remainders of 0 or 1 when divided by 3). So we've managed to show that $\gcd(\alpha^2 + \delta^2, \alpha^2 + 4\delta^2) = 1$. Combine this with the easy fact that $\gcd(\beta, \gamma) = 1$ and we see which parts of each side of (11) are relatively prime. Hence we must have

$$\beta^2 = \alpha^2 + (2\delta)^2, \tag{12}$$

$$\gamma^2 = \alpha^2 + \delta^2. \tag{13}$$

Applying the Pythagorean parametrization again to (12) we find $\alpha = \xi^2 - \eta^2$ and $\delta = \xi\eta$. Replacing these in (13) we get

$$(\xi^2 - \eta^2)^2 + (\xi\eta)^2 = \gamma^2.$$

This is of the same form as the original equation and we see that

$$\xi\eta = \delta < 2\gamma\delta = q < pq,$$

contradicting the initial claim that pq was minimal and proving that there are no solutions with p or q even.

We treat the remaining case that solutions p, q are both odd. Equation (8) now implies that

$$p^2 - q^2 = 2uv, \quad pq = u^2 - v^2,$$

provided $p \neq q$. Also one of u, v is necessarily even. Therefore

$$(u^2 - v^2)^2 + (uv)^2 = (pq)^2 + \frac{(p^2 - q^2)^2}{4} = \left(\frac{p^2 + q^2}{2}\right)^2,$$

which we have already seen is impossible. This completes the proof of the proposition. ■

So, if we look for a solution to (5) with positive distinct integers a, b, c and $a > c > 0$, say, then we must have $p = a + c = a - c = q$ implying that $c = 0$. Thus we have answered our original question.

PROPOSITION 3. *It is impossible for an arithmetic progression to have equal beginning, middle, and end sums.*

Four squares in arithmetic progression

Fermat wrote to Mersenne in May 1640 [8]. He included four challenges for Frenicle de Bessy, a number theorist in Paris:

Pour savoir si M. Frenicle ne procède point par tables, proposez lui de

(i) *Trouver un triangle rectangle duquel l'aire soit un nombre quarré;*

(ii) *Trouver deux quarréquarrés desquels la somme soit quarréquarrée;*

(iii) *Trouver quatre quarrés en proportion arithmétic continue;*

(iv) *Trouver deux cubes desquels la somme soit cube;*

> *S'il vous répond que jusques à un certain nombre de chiffres il a éprouvé que ces questions ne trouvent point de solution, assurez-vous qu'il procède par tables.*

From this we can detect Fermat's sensitivity to the difference between general proofs and empirical observations based on a table of factorizations of numbers (which today would be replaced by a computer search).

The first asks for a right-angled triangle (with integer length sides) whose area is a square. The Pythagorean parametrization reduces this to finding integer solutions for $x^4 - y^4 = z^2$, as shown by Pocklington [**4**, p. 615].

The fourth and second ask for solutions to $x^3 + y^3 = z^3$ and $x^4 + y^4 = z^4$. This was only the second time he had mentioned to his correspondents these cases of what became known as his Last Theorem. In about 1636, he sent Mersenne the same two problems and asked him to propose them to St. Croix. According to Dickson it was probably soon after, in 1637, that he made his famous note in the margin of his copy of Diophantus's *Arithmetica.*

The third challenge asks for four squares in arithmetic progression and this turns out to be related to our original question. Fermat seems to have been the first to look for such squares [**4**, p. 440]. That they do not exist follows from the fact that $x^4 - x^2y^2 + y^4 = z^2$ has only trivial solutions. We cannot be sure, but, as we shall discuss later, that might be what Fermat had in mind.

We can prove that four squares cannot be in arithmetic progression quite easily using Proposition 3 and the fact that the sum of the first n odd numbers is the nth square

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

(This has an easy geometric proof—can you find it? See Nelsen's proof without words [**12**] for something similar.) Thus if A^2, B^2, C^2, D^2 are four consecutive terms of an arithmetic progression with $0 < A < B < C < D$ we can take the sequence of consecutive odd numbers $2A + 1, 2A + 3, \ldots, 2D - 1$ and see that

$$\begin{aligned}(2A + 1) + (2A + 3) + \cdots + (2B - 1) &= (2B + 1) + (2B + 3) + \cdots + (2C - 1)\\ &= (2C + 1) + (2C + 3) + \cdots + (2D - 1),\end{aligned}$$

which contradicts Proposition 3.

By the same reasoning we cannot have four triangular numbers in arithmetic progression. More generally it follows from Proposition 3 that for any integers A, B, C, D and any real r we cannot have

$$A(A + r),\ B(B + r),\ C(C + r),\ D(D + r)$$

in arithmetic progression.

Euler's contribution

Euler proved in 1780 [**7**] that the product of four consecutive positive terms of an arithmetic progression cannot be a square. We will apply this result to find another proof of Proposition 3. Assume that we have an arithmetic progression with equal beginning, middle, and end sums. It leads to a solution of (6), which we may rewrite as

$$b(a - c)^2 = (a + c)(ac - b^2).$$

Letting $p = a + c, q = a - c$ as before, we find

$$4bq^2 = p(p^2 - q^2 - 4b^2)$$

and consequently

$$q^2(p + 4b) = (p - 2b)p(p + 2b).$$

In terms of a, b, c, this is

$$(a - c)^2(a + c + 4b) = (a + c - 2b)(a + c)(a + c + 2b).$$

If we multiply both sides of the above by $(a + c + 4b)$ we see that

$$(a + c - 2b)(a + c)(a + c + 2b)(a + c + 4b)$$

is a square. According to Euler this is impossible and we have a second proof of Proposition 3.

To close this circle of ideas we prove Euler's result. Suppose that there exist relatively prime integers $m, n \geq 1$ so that

$$m(m + n)(m + 2n)(m + 3n) = r^2. \qquad (14)$$

Where do the prime factors of r appear on the left-hand side of this equation? We must have $\gcd(m, m + 2n)$ dividing 2, $\gcd(m + n, m + 3n)$ dividing 2, and $\gcd(m, m + 3n)$ dividing 3, eight possibilities in all. This means that no prime bigger than 3 can appear in different terms of the factorization on the left. Thus, each of m, $m + n$, $m + 2n$, and $m + 3n$ is a square except for possible extra factors of 2 or 3. Checking the eight cases we see, for example, that $\{m, m + n, m + 2n, m + 3n\} = \{2A^2, B^2, 2C^2, D^2\}$ is not possible. This is because dividing $2A^2$, B^2, $2C^2$, and D^2 by 4 produces remainders 2, 1, 2, and 1 if each of A, B, C, and D are odd. But no arithmetic progression can have such remainders. One of A, B, C, and D may be even, but here too, checking each case, the remainders do not correspond to arithmetic progressions. It is routine to verify, modulo 3 and 4, that the only three possibilities for m, $m + n$, $m + 2n$, and $m + 3n$ are

(i) $\{A^2, B^2, C^2, D^2\}$
(ii) $\{6A^2, B^2, 2C^2, 3D^2\}$ or
(iii) $\{3A^2, 2B^2, C^2, 6D^2\}$

with A, B, C, and D relatively prime in pairs. We have already shown that (i) is impossible. We'll prove that (ii) cannot occur. Employ the easily verified identity

$$2\big(m(m + 2n) - (m + n)(m + 3n)\big) = m(m + n) - (m + 2n)(m + 3n)$$

from Pocklington [**15**] to get

$$4A^2C^2 - B^2D^2 = A^2B^2 - C^2D^2. \qquad (15)$$

Set

$$\alpha = 2AC, \quad \beta = BD, \quad \gamma = AB + CD, \quad \text{and} \quad \delta = AB - CD. \qquad (16)$$

Then we obtain $\alpha^2 - \beta^2 = \gamma\delta$ from (15) and $2\alpha\beta = \gamma^2 - \delta^2$ from (16). Therefore

$$(\alpha^2 - \beta^2)^2 + \alpha^2\beta^2 = \xi^2$$

for some ξ and by Proposition 2, we must have $\alpha = \beta$, which yields a contradiction. Part (iii) follows with an identical argument (as does part (i)) and this completes the proof. ■

Euler [**7**] used a slightly different approach. See also the discussion in Dickson [**4**, p. 635]. Interestingly, finding integer solutions to the general equation

$$m(m+n)(m+2n)\cdots\big(m+(k-1)n\big) = r^w, \tag{17}$$

(or showing they don't exist) has resisted many authors. The case with $n = 1$ has a long history, as described by Johnson [**10**], who gives relatively simple proofs of various cases. The question was eventually completely settled by Erdős and Selfridge [**6**] in a paper entitled "The product of consecutive integers is never a power." Recently Saradha [**16**] has shown that the only nontrivial solution to (17) (with $k \geq 3$ and $n \leq 22$ and $w = 2$) has $(m, n, k) = (18, 7, 3)$.

Back to Fermat's four challenges

Returning to Fermat's four challenges, we have seen that their impossibility follows, respectively, from the lack of nontrivial solutions to four Diophantine equations

(i) $x^4 - y^4 = z^2$,
(ii) $x^4 + y^4 = z^2$,
(iii) $x^4 - x^2y^2 + y^4 = z^2$,
(iv) $x^3 + y^3 = z^3$.

Frenicle did finally prove that $x^4 - y^4 = z^2$ has no nontrivial solutions with help from Fermat [**4**, p. 617]. He also came up with a formula supplying three squares in progression [**4**, p. 435]. But it fell to Euler to prove the impossibility of the first two cases of Fermat's Last Theorem [**4**, p. 545, p. 618] and that four squares cannot be in a progression, as we have seen [**7**].

Did Fermat himself have proofs? He certainly claimed that all four had only trivial solutions. We can only know with certainty that he had proved (i) and (ii). These two proofs, essentially identical, are rare examples of Fermat supplying his detailed arguments [**4**, p. 615], [**17**, p. 79]. In Weil's words, [**17**, p. 114]: "At that early date, Fermat had perhaps no more than plausibility arguments for the fact that these problems have no solution; but eventually he must have obtained a formal proof also for the third one, since we are told so by Billy in his *Inventum Novum*."

We cannot be sure what this formal proof of (iii) was since no trace of it appears in Fermat's writings. Weil laments that Billy did not find out more: "How grateful we should be to the good Jesuit, had he shown some curiosity toward such 'negative' statements ... "

One possibility is that Fermat worked directly with the equation $x^4 - x^2y^2 + y^4 = z^2$ and showed it has only trivial solutions using a proof like that of Proposition 2. This is appealing because the equations (i) to (iv) above are so similar.

A second possible approach, outlined by Weil and based on subsequent results of Euler that Fermat may have anticipated, is to work with the elliptic curve

$$y^2 = -x(x-1)(x-4). \tag{18}$$

It may be shown by the method of descent that this curve has only trivial rational solutions. This implies that four squares cannot be in arithmetic progression, as shown in [**17**, pp. 130–149]. It is an easy exercise to transform (14) into (18). This is done by Erdélyi [**5**].

A third approach, due to Erdélyi [**5**], is to rewrite (14) as

$$(m^2 + 3mn + n^2)^2 = r^2 + n^4. \tag{19}$$

He then shows, using the Pythagorean parametrization, that no solution in positive integers of (19) is possible, because each solution yields another that makes the quantity $(m + n)(m + 2n)$ smaller. This again is a classical proof by descent that Fermat could have used (he, of course, invented this technique). So there is no shortage of plausible ways Fermat could have proved this theorem.

As for the final challenge (iv), the proof of the impossibility of $x^3 + y^3 = z^3$ can be made to follow the same general lines but is harder than the others. It was probably not out of the reach of the "Prince of Amateurs" though; see the discussion in Mahoney's biography of Fermat [**11**, p. 357] and also Weil's thoughts [**17**, p. 118].

Arithmetic progressions with other ratios

We extend the discussion by letting $(S_1 : S_2 : S_3)$ denote the ratios of the sums (2). We have shown that $(1 : 1 : 1)$ is impossible. Here are some ratios involving the numbers 1, 2, 3 that are possible:

$$\begin{array}{ll}
(1:1:2) & 4, 5, 6;\ 7, 8;\ 9, 10, 11,\\
(1:1:3) & 1, 2;\ 3;\ 4, 5,\\
(1:2:2) & 6, 7, 8;\ 9, 10, 11, 12;\ 13, 14, 15,\\
(1:2:3) & 1;\ 2;\ 3,\\
(1:3:2) & 3;\ 4, 5;\ 6,\\
(1:3:3) & 2, 3;\ 4, 5, 6;\ 7, 8,\\
(2:1:3) & 12, 13, 14, 15, 16;\ 17, 18;\ 19, 20, 21, 22, 23.
\end{array} \tag{20}$$

Of course by changing the signs of each term in a sequence we can get the ratios in reversed order so that, for example, $-5, -4;\ -3;\ -2, -1$ yields $(3 : 1 : 1)$. As with Proposition 1 we may reduce the existence question to a Diophantine equation.

PROPOSITION 4. *There exists an arithmetic progression with three parts of a, b, and c terms and $(S_1 : S_2 : S_3) = (x : y : z)$ if and only if there exist positive integers a, b, c satisfying*

$$(xb - ya)c(b + c) + (zb - yc)a(a + b) = 0 \tag{21}$$

with the restriction that $xb \neq ya$ (or equivalently $zb \neq yc$).

We leave the proof to the reader. If this sequence exists and its terms differ by 1 then, as in (3), its first term e_1 must satisfy

$$2e_1 = \frac{ya^2 + xb^2}{ya - xb} - 2a + 1.$$

In the examples (20), we always have $a = c$. This is not a coincidence. When $a = c$, (21) reduces to $xb - ya = ya - zb$ or $2ya = (x + z)b$ and the restriction becomes $x \neq z$. This yields

PROPOSITION 5. *For positive integers x, y, z with $x \neq z$, there exists an arithmetic progression with three parts in ratio $(S_1 : S_2 : S_3) = (x : y : z)$.*

Proof. We may simply take $a = c = x + z$ and $b = 2y$. By Proposition 4 the desired progression exists completing the proof. ■

From this we obtain, for example,

$$(2:2:3) \quad 16, 17, 18, 19, 20;\ 21, 22, 23, 24;\ 25, 26, 27, 28, 29,$$

$$(2:3:3) \quad 20, 21, 22, 23, 24;\ 25, 26, 27, 28, 29, 30;\ 31, 32, 33, 34, 35.$$

Note that, since (21) is homogeneous in a, b, and c, any single solution yields an infinite family of solutions $\lambda a, \lambda b, \lambda c$ for λ a positive integer.

Next we look for progressions with ratios $(x : y : x)$. One way to solve (21) is to look for solutions of the form $c(b + c) = wa(a + b)$ and $xb - yc = w(ya - xb)$. From the first of these equations, let $c = a + b$ and $wa = b + c$. Therefore $a = 2$, $b = w - 1$, $c = w + 1$ and we require $w > 1$. This yields arithmetic progressions with ratios $(3w + 1 : w^2 - 1 : 3w + 1)$ parameterized by $w > 1$. This solution (when $w = 5$ and after multiplying by -2) gives

$$(2:3:2) \quad 3, 5, 7, 9, 11, 13;\ 15, 17, 19, 21;\ 23, 25.$$

A simpler example for this ratio is 1, 2, 3; 4, 5; 6. The remaining possibilities for ratios involving 1, 2, 3 are (1 : 2 : 1), (1 : 3 : 1), (2 : 1 : 2), (3 : 1 : 3), and (3 : 2 : 3). Solving (21) for b gives

$$b^2(xc + za) + b(za^2 + xc^2 - 2yac) - yac(a + c) = 0.$$

A necessary and sufficient condition for integer solutions is that the discriminant

$$z^2a^4 + x^2c^4 + \big(2(2y + x)(2y + z) - 4y^2\big)a^2c^2$$

be a square. Looking for the ratio (2 : 1 : 2), for example, we need $a^4 + 7a^2c^2 + c^4$ to be a square. Using the techniques of Proposition 2 it can be seen [**15, 2**] that this is impossible. Thus no arithmetic progression exists with beginning and end sums twice the middle sum. The other four possibilities are unresolved.

We finish with four challenges to the reader:

(i) For which values of x, y is $(x : y : x)$ a set of possible ratios for an arithmetic progression?

(ii) For positive integers x, y, z with $x \neq z$ is there a way to construct an arithmetic progression with the ratio $(x : y : z)$ and strictly positive terms? For example with $(x : y : z) = (3 : 2 : 1)$ Proposition 5 yields $-3;\ -2;\ -1$ but with more work we find

$$(3:2:1) \quad 9, 10, 11, \ldots, 24, 25, 26;\ 27, 28, 29, 30, 31, 32, 33;\ 34, 35, 36.$$

(iii) How many arithmetic progressions (with common difference, say, $\Delta = 1$ and parts of any size a, b, c but with $\gcd(a, b, c) = 1$) can represent a given ratio $(x : y : z)$?

(iv) When is it possible for the product of m consecutive terms of an arithmetic progression to be an nth power?

Acknowledgment. We thank the anonymous referees for their many improvements to the exposition and the additional reference.

REFERENCES

1. I. G. Bashmakova, *Diophantus and Diophantine Equations*, Mathematical Association of America, 1997.
2. J. H. E. Cohn, Squares in arithmetic progressions I, II, *Math. Scand.* **52**:1 (1983), 5–23.
3. J. de Billy, Inventum Novum, *Oeuvres de Fermat*, ed. C. Henry and P. Tannery, vol. III, 1894, pp. 325–398.
4. L. E. Dickson, *History of the Theory of Numbers*, vol. II, Chelsea, 1920.
5. T. Erdélyi, On the equation $a(a+d)(a+2d)(a+3d)=x^2$, *Amer. Math. Monthly* **107**:2 (2000), 166–169.
6. P. Erdős and J. L. Selfridge, The product of consecutive integers is never a power, *Illinois J. Math.* **19** (1975), 292–301.
7. L. Euler, *Mém. Acad. Sc. St. Petersbourg 8 années 1817–18*, Euler Opera Omnia, Ser. 1, ed. R. Fueter, vol. 5, 1944, pp. 48–60. (Latin)
8. P. de Fermat, *Oeuvres de Fermat*, ed. C. Henry and P. Tannery, vol. II, 1894, pp. 194–195.
9. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., Clarendon Press, 1979.
10. L. Johnson, On the Diophantine equation $x(x+1)\cdots(x+n-1)=y^k$, *Amer. Math. Monthly*, **47**:5 (1940), 280–289.
11. M. S. Mahoney, *The Mathematical Career of Pierre de Fermat*, 2nd ed., Princeton Univ. Press, 1994.
12. R. B. Nelsen, Proof without words: The cube as an arithmetic sum, this MAGAZINE **76**:2 (2003), 136.
13. *NY State Math. Teachers' Journal*, ed. M. C. Sachs, **48**:1 (1998), 56.
14. *NY State Math. Teachers' Journal*, ed. E. C. Wallace, **49**:1 (1999), 56.
15. H. C. Pocklington, Some Diophantine impossibilities, *Cambridge Phil. Soc.* **17** (1914), 108–121.
16. N. Saradha, Squares in products with terms in an arithmetic progression, *Acta Arithmetica* **86**:1 (1998), 27–43.
17. A. Weil, *Number Theory, An Approach through History*, Birkhauser, 1984.

From the *CUPM Curriculum Guide 2004—A Report by the Committee on the Undergraduate Program in Mathematics*

> Mathematics is universal: it underlies modern technology, informs public policy, plays an essential role in many disciplines, and enchants the mind.
>
> —from the *Introduction*

> Careful reasoning and communication are closely linked. A student who clearly understands a careful argument is capable of describing the argument to others. In addition, a requirement that students describe an argument or write it down tests whether understanding has truly occurred. All courses should include demands for students to speak and write mathematics, and more advanced courses should include more extensive demands. Communicating mathematical ideas with understanding and clarity is not only evidence of comprehension, it is essential for learning and using mathematics after graduation, whether in the workforce or in a graduate program.
>
> —from the section *Students majoring in the mathematical sciences*

The editor hopes that the mathematics offered in the MAGAZINE "enchants the mind" and that our mathematical communications (some written by students) stand up as good examples of this art.

NOTES

The Median Triangle in Hyperbolic Geometry

I. E. LEONARD
J. E. LEWIS
A. LIU
G. TOKARSKY
University of Alberta
Edmonton, Alberta
Canada T6G 2G1
isaac@cs.ualberta.ca

One of the problems assigned in our elementary Euclidean geometry course is to determine whether or not it is always possible to construct a triangle from the medians of an arbitrary triangle. Recall that a *median* is a line segment from a vertex to the midpoint of the opposite side. Such a triangle, if it exists, is called the *median triangle* of the original triangle. The problem is appropriate, because at that point of the course they know that it is not always possible to construct the *altitude triangle*, since the altitudes do not always satisfy the triangle inequality.

Some students approach this problem by drawing a particular triangle, constructing its medians, and then using the three medians to construct a triangle. The validity of their approach then depends on showing that the medians always satisfy the triangle inequality.

There is a way to circumvent having to prove the triangle inequality for the medians, as the following construction shows:

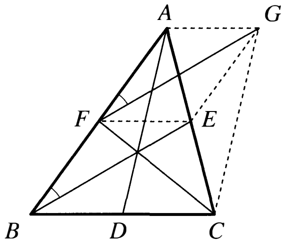


Figure 1 Constructing the median triangle

As in FIGURE 1, construct the medians AD, BE, and CF of triangle ABC. Construct the point G so that $\angle GFA = \angle EBF$ and $FG = BE$, then $BEGF$ is a parallelogram. Here we have used the theorem that a simple quadrilateral is a parallelogram if and only if one pair of opposite sides is congruent and parallel. Similarly, $EGAF$ and $ADCG$ are also parallelograms. Opposite sides of a parallelogram are the same length, so it follows that $GC = AD$, and that CFG is the median triangle. The construction works for any given triangle, so one can conclude that the median triangle always exists.

As an aside, note that FIGURE 1 shows how to construct from any given triangle GFC, another triangle ABC, of which the given triangle GFC is the median triangle.

Since E is the centroid of triangle GFC, it also can be used to show that the median triangle of the median triangle GFC is similar to the original triangle ABC, with proportionality 3 to 4.

Also, the proof that CFG is the median triangle uses the properties of a parallelogram, which in turn require the parallel postulate, and this places it clearly in the domain of Euclidean geometry. On the other hand, the construction itself does not require the parallel postulate. This note is concerned with the question as to whether this construction also produces a median triangle in hyperbolic geometry. In fact, we may ask if the median triangle even exists in hyperbolic geometry. The next section shows that it does, but the final section shows that the preceding construction never produces it.

Euclidean and noneuclidean geometries The distinction between Euclidean and noneuclidean geometries lies in their treatment of the parallel postulate. Euclid's famous fifth or parallel postulate states:

> *If a straight line falling on two straight lines makes the interior angles on the same side together less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are together less than two right angles.*

An equivalent form is attributed to the Scottish mathematician John Playfair and is known as Playfair's Axiom:

> *Through a given point not on a given line there can be drawn only one line parallel to the given line.*

In hyperbolic geometry, a type of noneuclidean geometry that retains the first four postulates of Euclid, the parallel postulate is replaced by the following:

> *Through a given point not on a given line there can be drawn at least two lines parallel to the given line.*

Hyperbolic geometry is logically consistent, and this discovery is sometimes credited as being the one that ushered in the modern era of mathematics. The development of hyperbolic geometry is attributed to Lobachevsky, Bolyai, and Gauss, but there is certainly some evidence that Euclid had an inkling of its existence. In *The Elements*, he carefully refrained from using the parallel postulate until the last possible moment. This fortunate occurrence means that the body of knowledge developed up to that point is valid in both Euclidean and hyperbolic geometry. This intersection is known as *absolute geometry*—it is the geometry obtained from the first four axioms of Euclidean geometry without the parallel postulate.

In absolute geometry, parallel lines do exist. Assuming that straight lines are infinite in extent, it can be proven without use of the parallel postulate that through a point not on a given line, it is possible to draw *at least one* line parallel to the given line. This is essentially Proposition 28 in the Elements, the last proposition prior to the introduction of the fifth postulate. Although it is not listed in Euclid, another consequence of the first four postulates is the following:

PROPOSITION. *In absolute geometry, the length of the median to any side of a triangle is less than the average of the lengths of the other two sides.*

Proof. In FIGURE 2, extend the median AD to G so that $AD = DG$.

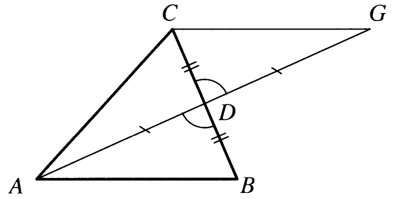


Figure 2 Twice a median is seen to be less than the sum of the other two sides

By side-angle-side congruency, triangles ABD and GCD are congruent, so $CG = AB$. The triangle inequality (Euclid's Proposition 20) gives $2 \cdot AD = AG < AC + CG$ so

$$AD < \frac{AC + AB}{2}. \qquad \blacksquare$$

It is now an easy matter to see that the median triangle exists in absolute geometry, In FIGURE 3, P is the point of intersection of the medians BE and CF.

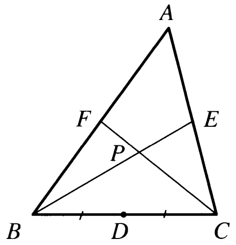


Figure 3 Existence of the median triangle in absolute geometry

From the proposition, the length of the median AD is less than $BF + CE$. But, by the triangle inequality

$$BF < BP + PF \quad \text{and}$$
$$CE < EP + PC,$$

so it follows that $AD < BE + CF$. This shows that the medians of ABC satisfy the triangle inequality, and so the median triangle exists.

The median triangle in hyperbolic geometry Many results of hyperbolic geometry seem strange to those familiar only with Euclidean geometry, the most striking one being that

1. *The sum of the interior angles of a triangle is strictly less than* 180°.

   An equivalent formulation is

2. *If $ABCD$ is a quadrilateral with $AD = BC$ and*

$$\angle BAD + \angle CBA = 180°,$$

   then $AB < CD$.

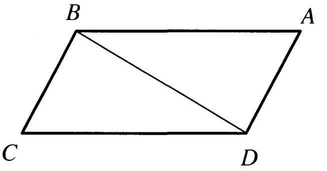


Figure 4 If $AD = BC$ and $\angle BAD + \angle CBA = 180°$, then $AB < CD$

To see why the two statements are equivalent, note that in FIGURE 4, the first statement implies that

$$\angle DAB + \angle ABD + \angle BDA < 180° = \angle DAB + \angle ABD + \angle DBC.$$

So, $\angle BDA < \angle DBC$, and it follows from the "Open Jaw Theorem" applied to triangles BDA and DBC that $AB < CD$.

Recall that the Open Jaw Theorem states that *If triangles ABC and $A'B'C'$ have $AB = A'B'$ and $AC = A'C'$, then $BC < B'C'$ if and only if $\angle A < \angle A'$.*

Moise [3, p. 121] calls this the "Hinge Theorem" and it is valid in absolute geometry.

Conversely, suppose we are given a triangle ABD. Let C be the point on the side of BD opposite to A such that $\angle CBA + \angle BAD = 180°$ and such that BC is congruent to AD. By the second statement, $AB < CD$, and so applying the Open Jaw Theorem we have

$$\angle BAD + \angle ADB + \angle DBA < \angle BAD + \angle CBD + \angle DBA = 180°.$$

One of the interesting consequences of the equivalence of these two statements is:

LEMMA. *In hyperbolic geometry, the line segment joining the midpoints of two sides of a triangle is less than half the length of the third side.*

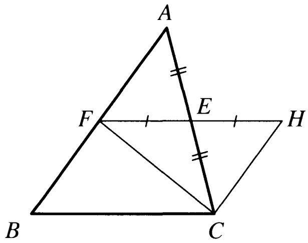


Figure 5 The segment joining midpoints

Proof. Extend the line through the midpoints E and F so that $EF = EH$. By side-angle-side congruence, triangles AFE and CHE are congruent, so $\angle AFE = \angle CHE$. Consequently, $BF = HC$ and

$$\angle BFH + \angle CHF = \angle BFH + \angle AFH = 180°,$$

and so by the previous discussion, $2FE = FH < BC$. ■

It now can be shown that the triangle GFC constructed in FIGURE 1 is *never* the median triangle in hyperbolic geometry. To see why, we have reproduced the construction in FIGURE 6 below, but have added the segment ED. Note that ED is not assumed to be collinear with EG, and recall that D, E, and F are the feet of the medians from A, B, and C.

The construction guarantees that GF is congruent to the median EB, but we will show that GC is never congruent to AD. For a contradiction, let us suppose that GC is congruent to AD.

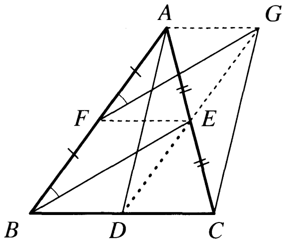


Figure 6 The original construction considered in hyperbolic geometry

From the lemma, we have

$$ED < \frac{1}{2}AB = FB.$$

In the quadrilateral $FBEG$ we have $\angle GFB + \angle EBF = 180°$, so

$$FB < EG.$$

Therefore $ED < EG$ and applying the Open Jaw Theorem to triangles GCE and DAE shows that $\angle GCE > \angle DAE$, that is, $\angle GCA > \angle DAC$.

On the other hand, the lemma also shows that

$$AG = EF < \frac{1}{2}BC = DC,$$

and applying the Open Jaw Theorem to triangles GCA and DAC shows that $\angle GCA < \angle DAC$, and this contradiction shows that GFC can never be the median triangle in hyperbolic geometry.

REFERENCES

1. Howard Eves, *A Survey of Geometry*, Allyn & Bacon Inc., Boston, 1972.
2. T. L. Heath, *Euclid: The Elements*, Dover Publications Inc., New York, 1956.
3. Edwin E. Moise, *Elementary Geometry from an Advanced Standpoint*, 3rd. ed., Addison Wesley Publishing Company, Inc., Reading, 1990.

Proof Without Words: The Sum of Cubes—An Extension of Archimedes' Sum of Squares

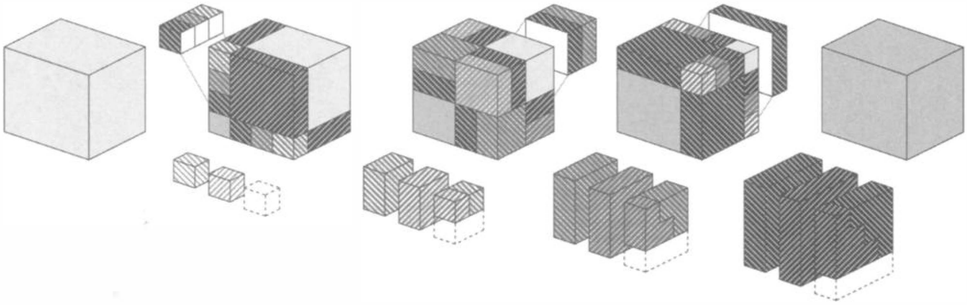
KATHERINE KANIM
New Mexico State University
Las Cruces, NM 88003
kkanim@nmsu.edu

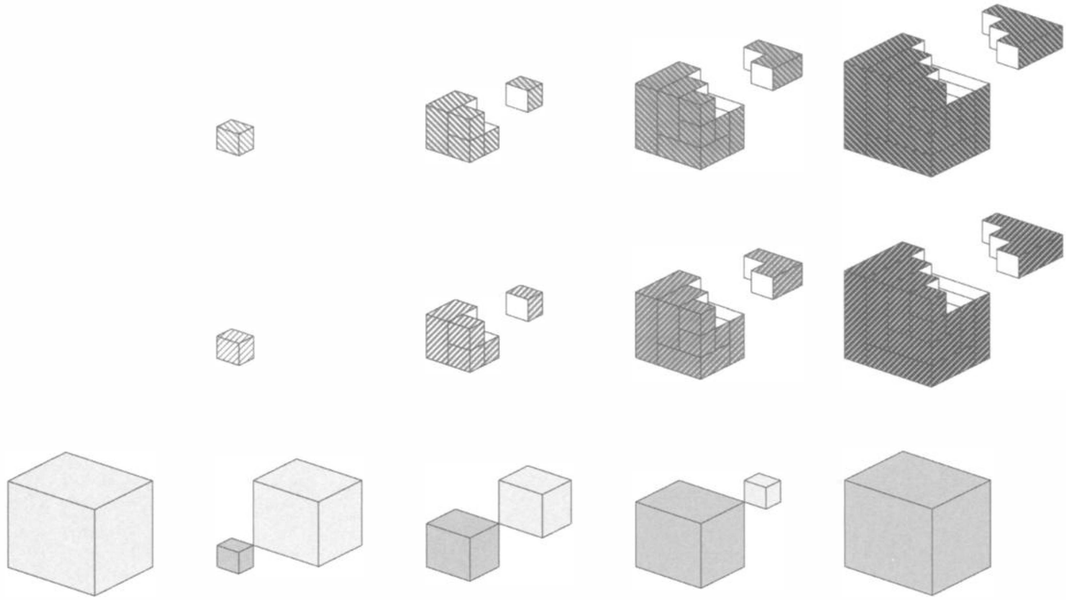
Archimedes' proposition determining a sum of squares has been displayed geometrically in a previous Proof Without Words [1]. We generalize to a sum of cubes, stating the result first as Archimedes might have, in the geometric language he used for his sum of squares. Then we give a proof befitting the geometric language.

> If a series of any number of lines be given, which exceed one another by an equal amount, and the difference be equal to the least, and if other lines be given equal in number to these and in quantity to the greatest, the cubes on the lines equal to the greatest, plus the cube on the greatest and the triplicate of the rectangular solids contained by the least and the squares of the lines exceeding one another by an equal amount, less the rectangular solids contained by the square on the least and those exceeding one another by an equal amount, will be the quadruple of all the cubes on the lines exceeding one another by an equal amount.

In modern symbolism,

$$(n+1)n^3 + 3\sum_{i=1}^{n} i^2 - \sum_{i=1}^{n} i = 4\sum_{i=1}^{n} i^3.$$





Further investigation has revealed a generalization for arbitrary exponents. A geometric figure for a sum of fourth powers, which proves

$$(n+1)n^4 + 6\sum_{i=1}^{n} i^3 - 4\sum_{i=1}^{n} i^2 + \sum_{i=1}^{n} i = 5\sum_{i=1}^{n} i^4,$$

is challenging, but is underway via a diagram involving shadows of hypercubes.

Acknowledgment. Thank you to David Pengelley for suggestions and support, and to Roger Nelsen for the beautiful electronic layout of the figure.

REFERENCES

1. K. Kanim, Proof without words: How did Archimedes sum squares in the sand?, this MAGAZINE **74** (2001), 314–315.

An October Warning

Possibly Gilman ought not to have studied so hard. Non-Euclidean calculus and quantum physics are enough to stretch any brain; and when one mixes them with folklore, and tries to trace a strange background of multi-dimensional reality behind the ghoulish hints of the Gothic tales and the wild whispers of the chimney-corner, one can hardly expect to be wholly free from mental tension.... The professors at Miskatonic had urged him to slacken up.... But all these precautions came late in the day, so that Gilman had some terrible hints from the dreaded *Necronomicon*....

From *The Dreams in the Witch-House*, by H. P. Lovecraft

A (Not So) Complex Solution to $a^2 + b^2 = c^n$

ARNOLD M. ADELBERG
Grinnell College
Grinnell, IA 50112
adelbe@math.grinnell.edu

ARTHUR T. BENJAMIN
Harvey Mudd College
Claremont, CA 91711
benjamin@math.hmc.edu

DAVID I. RUDEL
Dartmouth College
Hanover, NH 03755
david.rudel@dartmouth.edu

Everyone knows how easy it is to describe all solutions to the Diophantine equation $a^2 + b^2 = c^2$, and how difficult it is to prove the nonexistence of solutions to $a^n + b^n = c^n$ for $n > 2$. Mixing the easy equation with the hard one, we shall demonstrate that for $n \geq 2$, all solutions to $a^2 + b^2 = c^n$ can also be easily obtained by elementary number theory. We use only the simplest properties of the *Gaussian integers*, the complex numbers whose real and imaginary parts are both integers.

To set the stage, recall the situation when $n = 2$. The Pythagorean triples of positive integers that are primitive (that is, have no common prime factors) all have the form (a, b, c) or (b, a, c) where $a = x^2 - y^2$, $b = 2xy$, $c = x^2 + y^2$, where $x > y$ are relatively prime integers of opposite parity. This can be expressed more compactly using the Gaussian integer $z = x + yi$ and its conjugate $\overline{z} = x - yi$, whereby

$$a = \mathrm{Re}(z^2), \qquad b = \mathrm{Im}(z^2), \qquad c = \overline{z}z.$$

All solutions to $a^2 + b^2 = c^2$ are multiples of the primitive solutions.

In general, for $n \geq 2$, we demonstrate that all primitive solutions of $a^2 + b^2 = c^n$ have the form (a, b, c), where

$$\begin{array}{llll} a = \mathrm{Re}(z^n), & b = \mathrm{Im}(z^n), & c = \overline{z}z & \text{when } n \text{ is odd,} \\ a = \mathrm{Re}(z^n\omega), & b = \mathrm{Im}(z^n\omega), & c = \pm\overline{z}z & \text{when } n \text{ is even,} \end{array}$$

with $z = x + yi$, where x and y are relatively prime integers of opposite parity. In the even case, ω comes from the set of units $\{1, i, -1, -i\}$.

If $n > 2$, *not* all solutions to $a^2 + b^2 = c^n$ are multiples of primitive ones. The primitive solutions may be deduced by classical methods or from the following two theorems.

THEOREM 1. *When $n \geq 3$ is odd, integers a, b, and c satisfy $a^2 + b^2 = c^n$ if and only if*

$$a = \mathrm{Re}(z), \qquad b = \mathrm{Im}(z), \qquad c = \prod_{t=0}^{(n-1)/2} \overline{z}_t z_t, \tag{1}$$

where each z_t is a Gaussian integer and $z = \prod_{t=0}^{(n-1)/2} \overline{z}_t^t z_t^{n-t}$.

Proof. It is easy to see that (1) will generate solutions to $a^2 + b^2 = c^n$, since

$$a^2 + b^2 = \overline{z}z = \prod_{t=0}^{(n-1)/2} \overline{z}_t^n z_t^n = c^n.$$

To prove the other direction suppose that $a^2 + b^2 = c^n$, where n is an odd positive integer. We will use elementary number theory to prove that (1) is necessary for solutions. Let c^n have prime factorization

$$c^n = \prod_{j=1}^{k} p_j^{\alpha_j} \prod_{i=1}^{m} q_i^{\beta_i},$$

where $p_1 = 2$, $p_j \equiv 1 \pmod 4$, $j = 2 \ldots, k$, and $q_i \equiv 3 \pmod 4$, $i = 1 \ldots m$. We assume the primes are distinct, and hence $n \mid \alpha_j$ and $n \mid \beta_i$ for all i and j. Further, since c^n is the sum of two squares, it is well known [**2**] that $2 \mid \beta_i$ and since n is odd, $2n \mid \beta_i$ for all i.

Next we factor c^n into Gaussian primes, which consist of the traditional primes $q \equiv 3 \pmod 4$ and Gaussian integers $w = u + vi$ that satisfy $u^2 + v^2 = p$, where p is a prime not congruent to $3 \pmod 4$ [**1**]. By standard number theory [**2**], every such prime p is the sum of two squares. That is, for $j = 1 \ldots, k$, each p_j above can be written as $p_j = \overline{\rho}_j \rho_j$ where ρ_j is a Gaussian prime.

Summarizing, c^n has Gaussian prime factorization

$$(a + bi)(a - bi) = a^2 + b^2 = c^n = \prod_{j=1}^{k} (\overline{\rho}_j \rho_j)^{\alpha_j} \prod_{i=1}^{m} q_i^{\beta_i},$$

where $n \mid \alpha_j$, $j = 1 \ldots, k$, and $2n \mid \beta_i$, $i = 1 \ldots, m$. By the unique factorization of Gaussian integers (up to multiplication by units) into Gaussian primes, we must have

$$a + bi = \prod_{j=1}^{k} \overline{\rho}_j^{\gamma_j} \rho_j^{\delta_j} \prod_{i=1}^{m} q_i^{\beta_i/2} \omega, \tag{2}$$

and

$$a - bi = \prod_{j=1}^{k} \overline{\rho}_j^{\delta_j} \rho_j^{\gamma_j} \prod_{i=1}^{m} q_i^{\beta_i/2} \overline{\omega},$$

where ω is a unit, and for $j = 1 \ldots k$, we have $\gamma_j, \delta_j \geq 0$, and $\gamma_j + \delta_j = \alpha_j$.

Now define $r_j = \gamma_j \bmod n$ for $j = 1 \ldots, n$. Since $\gamma_j + \delta_j = \alpha_j$ is a multiple of n, we may write $\gamma_j = ns_j + r_j$ and $\delta_j = nt_j + (n - r_j)$, where $s_j, t_j \geq 0$ and $0 \leq r_j < n$. Hence, (2) becomes

$$a + bi = \prod_{j=1}^{k} (\overline{\rho}_j^{s_j} \rho_j^{t_j})^n \overline{\rho}_j^{r_j} \rho_j^{n-r_j} \left(\prod_{i=1}^{m} q_i^{\beta_i/2n} \right)^n \omega. \tag{3}$$

For a given exponent $0 \leq e \leq n$, the product of numbers of the form $\overline{w}^e w^{n-e}$ will still be of that form, and replacing w by $\overline{w}$ if necessary, we can assume $e \leq (n-1)/2$. Hence for $t = 0, 1, \ldots, (n-1)/2$, we let z_t denote the product of all terms in (3) of the form $\overline{w}^t w^{n-t}$. Note that terms of the form w^n have the form $\overline{w}^0 \overline{\overline{w}}^n$, and that ω is itself an nth (odd) power. Hence

$$a + bi = \prod_{t=0}^{(n-1)/2} \overline{z}_t^t z_t^{n-t},$$

and (1) follows. ■

THEOREM 2. *When $n \geq 2$ is even, integers a, b, and c satisfy $a^2 + b^2 = c^n$ if and only if*

$$a = r^{n/2}\,\mathrm{Re}(z\omega), \qquad b = r^{n/2}\,\mathrm{Im}(z\omega), \qquad c = \pm r \prod_{t=0}^{(n-2)/2} \bar{z}_t z_t \tag{4}$$

where r is a positive integer, ω is a unit, each z_t is a Gaussian integer, and $z = \prod_{t=0}^{(n-2)/2} \bar{z}_t^{t} z_t^{n-t}$.

Proof. The proof follows along the same lines as the previous one. The only subtlety to point out is that although 2 and n divide β_i, $2n$ might not divide β_i. However, the term $\prod_{i=1}^m q_i^{\beta_i/2} = (\prod_{i=1}^m q_i^{\beta_i/n})^{n/2}$ and the (integer) term of the form $\bar{z}_t^{n/2} z_t^{n/2}$ can be absorbed into the integer $r^{n/2}$. ■

Acknowledgment. The authors gratefully acknowledge the assistance of Reba Schuller and the referees for helpful suggestions.

REFERENCES

1. David M. Burton, *Elementary Number Theory*, Allyn and Bacon, Inc., Boston, 1980.
2. R. M. Young, *Excursions in Calculus: An Interplay of the Continuous and the Discrete*, Dolciani Math. Exp. 13., MAA, Washington, D.C., 1992.

On the Two-Box Paradox

ROBERT A. AGNEW
Discover Financial Services
Riverwoods, IL 60015-3851
robertagnew@discoverfinancial.com

On a game show, you are presented with two identical boxes. Both boxes contain positive monetary prizes, one twice the other. You are allowed to pick one box and observe the prize $x > 0$, after which you can choose to trade boxes. In terms of simple expected value, it is *always* better to trade since $\frac{1}{2}(2x) + \frac{1}{2}\left(\frac{x}{2}\right) = \frac{5x}{4} > x$. That is the paradox.

Simple thought experiments suggest that a sufficiently large observed prize would cause a player not to trade, despite the mathematical computation of expected value. In individual cases, this creates some threshold, which depends on the observed prize, for ceasing to trade. A player may have in mind prior probabilities about what prizes the game show would offer, so that an observed prize of $10,000, for instance, would not yield equal *judgmental* odds of $20,000 or $5,000 in the unobserved box. The judgmental probability approach to the two-box problem seeks to develop optimal threshold strategies in terms of prior distributions on the set of possible prizes. Recent articles in this MAGAZINE have focused on the judgmental probability approach, although they have also discussed the second line of attack on this problem, expected utility [**2**, **3**].

In expected utility theory, it is assumed than an individual has an underlying utility function for wealth. This utility function is increasing because it is presumed that an individual will always prefer more wealth to less wealth. In addition, the utility function is concave because it is presumed that an individual will have nonincreasing marginal utility for wealth. The utility function u is thus an increasing, concave function from the positive half line into the real line. The scaling on this function is unimportant because a positive linear transformation $a + bu$, with $b > 0$, is equivalent for individual decision making. Finally, linear utility $u(w) = w$, which is inherent in the simple statement of the two-box paradox, really represents a boundary case. Economists normally assume strictly diminishing marginal utility for wealth, for instance, a person prefers \$100,000 of wealth to \$50,000 but he has less use for an *additional* dollar when he has \$100,000 than when he has \$50,000.

To recapitulate, we assume that an individual has an underlying utility function, or preference function, for wealth, even though he or she may not have detailed this function. The expected utility hypothesis takes this notion of utility of wealth one step further. It is assumed that a rational individual will explicitly model his utility function for wealth u and will select among risky prospects based on maximal expected utility. In the context of the two-box game, it is assumed that a rational individual will trade if and only if $(1/2)u(w_0 + 2x) + (1/2)u(w_0 + x/2) > u(w_0 + x)$, where w_0 is his initial wealth and x is the observed prize. Note that we are not deviating from the 50-50 prize distribution between $2x$ and $x/2$, which is inherent in the statement of the two-box game. There is no injection of judgmental probabilities. Our subsequent analysis is based on expected utility theory alone.

Expected utility has a rich history dating back to Daniel Bernoulli's 1738 resolution of the St. Petersburg paradox. In the St. Petersburg game, a fair coin is flipped until a head occurs, and a player receives 2^n when the first head occurs at trial n. It is easy to see that expected payoff is infinite. Nevertheless, Bernoulli observed that no rational individual would pay a huge amount to play this game and he resolved the paradox by assuming logarithmic utility for wealth. Later, Karl Menger in 1934 observed that full resolution of the St. Petersburg paradox requires a utility function that is *bounded above*. A nice historical perspective is provided by Fonseca and Ussher [**4**].

Once again, this note focuses on a pure expected utility approach, without reference to any prior distribution; the only probabilistic element is the 50-50 prize distribution between $2x$ and $x/2$, which is associated with the trading gamble when a player in the two-box game is faced with two identical boxes and an observed prize of x. In that context, we show that the two-box paradox is confined to unbounded utility functions and that common bounded utility functions have well-defined optimal threshold strategies.

Utility functions and risk Assume that an individual has utility $u(w)$ for wealth $w > 0$. Economists generally assume that rationality requires the utility function u to be strictly increasing and concave ($u' > 0$ and $u'' \le 0$ assuming differentiability). Normally, the utility function should be strictly concave to reflect a person's diminishing preference for an *additional* dollar at increasing wealth levels. Economists also generally postulate that a rational individual facing alternative risky prospects, or gambles, will choose to maximize expected utility.

Beyond these general notions of rationality, economists have defined local measures of absolute risk aversion as

$$-u''(w)/u'(w)$$

and relative risk aversion as

$$-wu''(w)/u'(w).$$

These measures follow naturally from Taylor approximations to the so-called *risk premia* associated with small random perturbations to wealth, either additive or proportional [**4, 5, 6**]. Given utility function u and wealth level w, the risk premium π associated with an additive random wealth perturbation ϵ is defined by the functional equation $u(w - \pi) = \mathrm{E}(u(w + \epsilon))$, where π is the premium that one is willing to pay to avoid the random wealth perturbation. We have

$$u(w-\pi) \cong u(w) - u'(w)\pi \quad \text{and} \quad u(w+\epsilon) \cong u(w) + u'(w)\epsilon + u''(w)\epsilon^2/2.$$

If $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2$, then

$$\mathrm{E}\left(u(w+\epsilon)\right) \cong u(w) + u''(w)\sigma^2/2 \quad \text{and hence} \quad \pi \cong \left(- u''(w)/u'(w)\right)\sigma^2/2,$$

with units $\$ = (1/\$) \times \$^2$. Relative risk aversion has a similar interpretation with respect to the proportional risk premium defined by the functional equation

$$u\big((1-\pi)\cdot w\big) = \mathrm{E}\Big(u\big((1+\epsilon)\cdot w\big)\Big),$$

that is, π is the proportion of wealth one will pay to avoid a proportional wealth perturbation ϵ. In this case, we have dimensionless $\pi \cong (-wu''(w)/u'(w))\sigma^2/2$.

Up to simple scaling, identically zero risk aversion, either absolute or relative, implies a linear utility function $u(w) = w$. Constant *absolute* risk aversion $\alpha > 0$ implies a utility function of the form $u(w) = -e^{-\alpha w}$, which is bounded above. Once again, scaling is unimportant and the utility function $u(w) = 100(1 - e^{-\alpha w})$, which takes positive values, is completely equivalent for our purposes. The higher the risk aversion parameter α, the less tolerance the individual has for additive wealth perturbation at any level of wealth.

Up to simple scaling, constant positive *relative* risk aversion $\beta > 0$ implies a utility function of the form $u(w) = w^{1-\beta}$ for $\beta \in (0,1)$, $u(w) = \ln(w)$ for $\beta = 1$, or $u(w) = -w^{1-\beta}$ for $\beta > 1$, where the latter function is bounded above. Once again, scaling is unimportant but we can't avoid negative utility values for small levels of wealth when $\beta \geq 1$. The higher the risk aversion parameter β, the less tolerance the individual has for proportional wealth perturbation at any level of wealth.

Someone with a linear utility function is entirely indifferent to risk in that doubling his fortune doubles his satisfaction, and of course he is subject to the two-box paradox. The other cases aren't so obvious and we deal with them in the next two sections. We don't claim that these constant risk aversion utility functions are the only ones worth considering, although they have been much discussed. Some economists have suggested that absolute risk aversion should decrease with wealth while relative risk aversion should increase [**5**].

Two-box paradox and unbounded utility Denote initial wealth by w_0 and the observed prize by x. The two-box paradox arises when an individual prefers to trade, or gamble, without regard to his initial wealth position or the observed prize. In mathematical terms, a utility function u is subject to the paradox if

$$\frac{1}{2}u(w_0 + 2x) + \frac{1}{2}u\left(w_0 + \frac{x}{2}\right) > u(w_0 + x)$$

or equivalently

$$u(w_0 + 2x) - u(w_0 + x) > u(w_0 + x) - u(w_0 + x/2)$$

for any $w_0, x > 0$.

For an individual subject to the paradox, the gain in satisfaction from doubling the observed prize (for instance, \$10,000 to \$20,000) always exceeds the loss in satisfaction from halving the observed prize (for instance, \$10,000 to \$5,000), regardless of his initial wealth position. The following result shows that this condition is not confined to linear utility.

PROPOSITION 1. *If* $u(w) = w^\gamma$ *for* $\gamma \in (0,1]$ *or* $u(w) = \ln(w)$, *then the paradox occurs.*

Proof. For any $w_0, x > 0$,

$$\frac{(w_0 + 2x)^\gamma + (w_0 + x/2)^\gamma}{2} - (w_0 + x)^\gamma > (w_0 + x)^\gamma \left(\frac{t^\gamma + t^{-\gamma}}{2} - 1\right)$$
$$= (w_0 + x)^\gamma (t^\gamma - 1)(1 - t^{-\gamma})/2 > 0,$$

where $t = (w_0 + 2x)/(w_0 + x) > (w_0 + x)/(w_0 + x/2) > 1$. Moreover,

$$\frac{\ln(w_0 + 2x) + \ln(w_0 + x/2)}{2} - \ln(w_0 + x) = \ln\left(\frac{\sqrt{w_0^2 + 5w_0x/2 + x^2}}{w_0 + x}\right)$$
$$> \ln\left(\frac{\sqrt{w_0^2 + 2w_0x + x^2}}{w_0 + x}\right) = 0. \quad \blacksquare$$

We have just shown that conventional unbounded utility functions are subject to the paradox. We next show that the paradox occurs only for utility functions that are unbounded above. This result has already been proved [**3**], but our proof avoids any reference to prior distributions.

PROPOSITION 2. (BRAMS AND KILGOUR) *A necessary condition for the paradox is that the utility function u be unbounded above.*

Proof. Let $a_n = u(w_0 + 2^{n+1}) - u(w_0 + 2^n)$ for $n \geq 0$, and let $s_n = \sum_{k=0}^n a_k = u(w_0 + 2^{n+1}) - u(w_0 + 1)$. If the paradox arises, then a_n is a positive, increasing sequence so that $s_n \uparrow \infty$, and thus u is unbounded above. $\blacksquare$

We conclude that utility functions that are bounded above have at least the potential for simple threshold strategies for ceasing to trade in the two-box game.

Bounded utility We now show that simple threshold strategies exist for two families of bounded utilities. For the constant absolute risk aversion family, the prize threshold is independent of initial wealth. For the family of functions with constant relative risk aversion, the prize threshold is proportional to initial wealth. These results are not surprising. Constant absolute risk aversion implies the same sensitivity to additive wealth perturbation across the spectrum of existing wealth. Constant relative risk aversion, on the other hand, implies the same sensitivity to proportional wealth perturbation across the spectrum of existing wealth. Once again, the optimal prize thresholds are independent of utility function scaling.

PROPOSITION 3. *Let $u(w) = -e^{-\alpha w}$ for some $\alpha > 0$, so that absolute risk aversion is constant and positive. Then the optimal threshold strategy in the two-box game is to trade when the observed prize x is less than $x^\star = -2\ln(\theta)/\alpha$, where $\theta = (\sqrt{5} - 1)/2$, and not to trade when $x \geq x^\star$. The optimal threshold $x^\star$ is independent of initial wealth w_0.*

Proof. For $w_0, x > 0$,

$$\frac{1}{2}u(w_0 + 2x) + \frac{1}{2}u(w_0 + x/2) - u(w_0 + x) = -\frac{1}{2}e^{-\alpha(w_0+x/2)}f(t),$$

where $t = e^{-\alpha x/2}$ and $f(t) = t^3 + 1 - 2t$ (graphed in FIGURE 1).

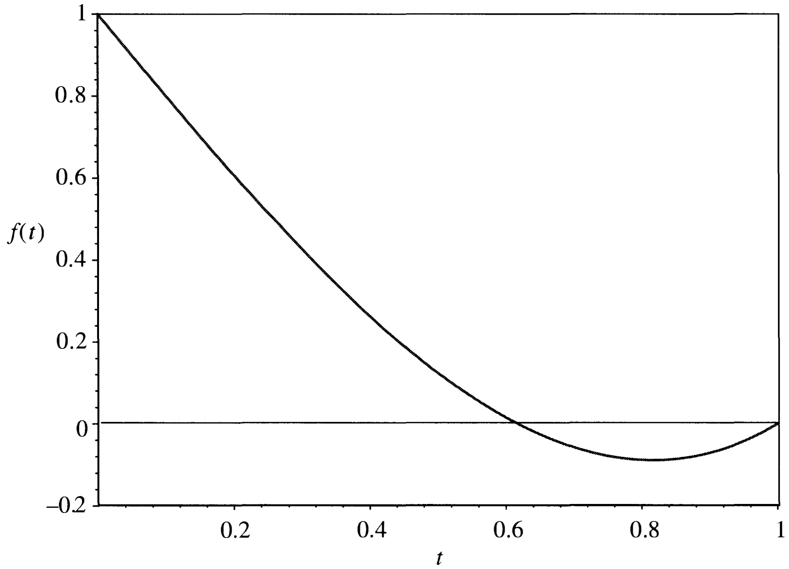


Figure 1 The function $f(t) = t^3 + 1 - 2t$

It is easy to check that f is strictly convex $(f'' > 0)$ on $(0, 1)$ with $f(0) = 1$, $f(1) = 0$, $f'(0) < 0$, and $f'(1) > 0$. The relation $\theta^2 = 1 - \theta$ implies that $f(\theta) = 0$ and we conclude that θ is the unique zero of f in $(0, 1)$, with $f(t) > 0$ for $t \in (0, \theta)$ and $f(t) < 0$ for $t \in (\theta, 1)$. Since $-(1/2)e^{-\alpha(w_0+x/2)}f(t)$ is positive if and only if $t > \theta$, which is the same as $x < x^\star$, it follows that $x^\star$ is the optimal threshold point, independent of w_0. ■

PROPOSITION 4. *Let $u(w) = -w^{-\gamma}$ for some $\gamma > 0$ so that relative risk aversion $\beta = \gamma + 1$ is constant and greater than one. Then, the optimal threshold strategy in the two-box game is to trade when the observed prize $x < x^\star = w_0\phi/(1 - \phi)$, where w_0 is initial wealth and ϕ is the unique root in $(0, 1)$ of $(1 + t)^{-\gamma} + (1 - t/2)^{-\gamma} = 2$, and not to trade when $x \geq x^\star$. The optimal threshold $x^\star$ is proportional to initial wealth w_0.*

Proof. For $w_0, x > 0$,

$$\frac{1}{2}u(w_0 + 2x) + \frac{1}{2}u(w_0 + x/2) - u(w_0 + x) = -\frac{1}{2}(w_0 + x)^{-\gamma} f(t),$$

where this time $t = x/(w_0 + x)$ and $f(t) = (1 + t)^{-\gamma} + (1 - t/2)^{-\gamma} - 2$ (graphed in FIGURE 2 for $\gamma = 1$).

It is easy to check that f is strictly convex $(f'' > 0)$ on $(0, 1)$ with $f(0) = 0$, $f(1) = 2^{-\gamma} + 2^{\gamma} - 2 = (2^{\gamma} - 1)(1 - 2^{-\gamma}) > 0$, $f'(0) < 0$, and $f'(1) > 0$. We conclude that there exists a unique zero ϕ of f in $(0, 1)$ with $f(t) < 0$ for $t \in (0, \phi)$ and $f(t) > 0$ for $t \in (\phi, 1)$. Since $-(1/2)(w_0 + x)^{-\gamma}f(t)$ is positive if and only if $t < \phi$, which is the same as $x < x^\star$, it follows that $x^\star$ is the optimal threshold point and that it is proportional to w_0. ■

We remark that an extension of Proposition 4 is easily obtained for a utility function of the form $u(w) = -(\eta + w)^{-\gamma}$ where $\gamma > 0$ and $\eta > 0$. The proof goes through the same way with optimal threshold $x^\star = (\eta + w_0)\phi/(1 - \phi)$. This utility function, which falls into the category that Gollier **[5]** calls *harmonic absolute risk aversion*, exhibits decreasing absolute risk aversion $(\gamma + 1)/(\eta + w)$ and increasing relative risk aversion $(\gamma + 1)w/(\eta + w)$, although both types of risk aversion are reduced by in-

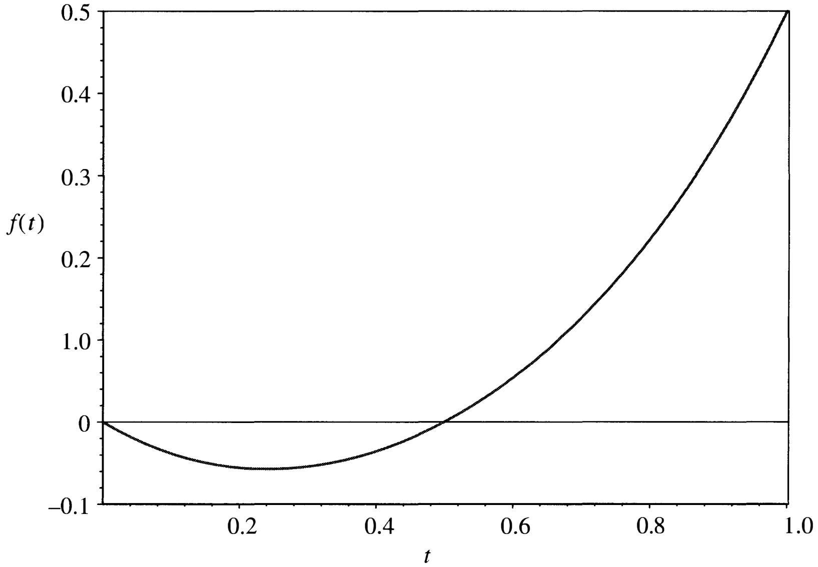


Figure 2 The function $f(t) = (1 + t)^{-\gamma} + (1 - t/2)^{-\gamma} - 2$

clusion of the additional parameter. The following two examples illustrate further the distinct behavioral differences that are implied by the two families of utility functions in Propositions 3 and 4, constant absolute risk aversion and constant relative risk aversion.

EXAMPLE 1. *Suppose an individual has utility function* $u(w) = 100(1 - e^{-.0001w})$, *which exhibits constant absolute risk aversion with parameter* $\alpha = .0001$. *Then, he will trade if and only if the observed prize is less than $9,624.24, regardless of his initial wealth. If this person observes a prize of $5,000, he will trade, no matter whether his existing wealth is $10 or $1,000,000. If, on the other hand, he observes a prize of $10,000, he will not trade under any circumstance.*

EXAMPLE 2. *Suppose an individual has utility function* $u(w) = 100 - 10000w^{-1}$, *which exhibits constant relative risk aversion with parameter* $\beta = 2$. *Then, he will trade if and only if the observed prize is less than his initial wealth since the root* $\phi = 1/2$. *If this person observes a prize of $10,000, he will trade if his existing wealth is less than $10,000 but he will not trade if his existing wealth is $10,000 or greater.*

In conclusion, we have demonstrated that the two-box paradox, like the St. Petersburg paradox, can be resolved by bounded utility of wealth and that for traditional bounded utility functions, simple threshold strategies are optimal.

REFERENCES

1. R. A. Agnew, Inequalities with application in economic risk analysis, *J. Applied Probability* **9** (1972), 441–444.
2. N. M. Blachman and D. M. Kilgour, Elusive optimality in the box problem, this MAGAZINE **74** (2001), 171–181.
3. S. J. Brams and D. M. Kilgour, The box problem: to switch or not to switch, this MAGAZINE **68** (1995), 27–34.
4. G. L. Fonseca and L. J. Ussher, Choice under risk and uncertainty, *The History of Economic Thought Website, Department of Economics of the New School for Social Research*, `http://cepa.newschool.edu/het/essays/uncert/choicecont.htm`.

5. C. Gollier, *The Economics of Risk and Time*, MIT Press, Cambridge, Mass., 2001.
6. J. W. Pratt, Risk aversion in the small and in the large, *Econometrica* **32** (1964), 122–136.

A Conceptual Proof of Cramer's Rule

RICHARD EHRENBORG
University of Kentucky
Lexington, KY 40506-0027
jrge@ms.uky.edu

THEOREM. (CRAMER'S RULE) *Let A be an invertible $n \times n$ matrix. Then the solutions x_i to the system $A\mathbf{x} = \mathbf{b}$ are given by*

$$x_i = \frac{\det(A_i)}{\det(A)}, \tag{1}$$

where A_i is the matrix obtained from A by replacing the ith column of A by $\mathbf{b}$.

Proof. The classical way to solve a linear equation system is by performing row operations: (i) add one row to another row, (ii) multiply a row with a nonzero scalar and (iii) exchange two rows. We show that the quotient in equation (1) will not change under row operations.

Under the first row operation, the values of the two determinants $\det(A_i)$ and $\det(A)$ will not change, since determinants are invariant under this row operation. Under the second row operation both determinants will gain the same factor, which cancels in the quotient. Finally, under the third row operation both determinants will switch sign, which again cancels in the quotient.

Since every invertible matrix A can be row reduced to the identity matrix, it is now enough to prove Cramer's rule for the identity matrix. However, this is a straightforward task. ■

A Parent of Binet's Formula?

B. SURY
Stat-Math Unit
Indian Statistical Institute
8th Mile Mysore Road
Bangalore 560 059 India
sury@isibang.ac.in

The famous Binet formula for the Fibonacci sequence $F_1 = 1 = F_2$, $F_{n+2} = F_n + F_{n+1}$ is the identity

$$F_n = \frac{\phi^n - (-1/\phi)^n}{\sqrt{5}},$$

where ϕ is the golden ratio $(1 + \sqrt{5})/2$.

As we all know, many identities—even quite complicated ones—once written down, can be verified by anybody who can perform elementary algebraic manipulations. However, discovering it may not be easy at all. Binet's formula too can be verified easily. As for arriving at it, one method is to look for exponential solutions to the difference equation that defines the Fibonacci numbers. Here is another way to arrive at Binet's formula by producing a polynomial identity that perhaps could be regarded as a parent of Binet's formula.

Note first that the golden ratio ϕ satisfies the identities

$$\phi - \frac{1}{\phi} = 1, \quad \phi + \frac{1}{\phi} = \sqrt{5}.$$

Let us look at the polynomial

$$F_n(X, Y) := \sum_{i=0}^{\lfloor n/2 \rfloor} (-1)^i \binom{n-i}{i} X^i Y^i (Y + X)^{n-2i}.$$

It is an easy exercise in induction on n to show that

$$F_n(X, Y) = X^n + X^{n-1}Y + \cdots + XY^{n-1} + Y^n.$$

Indeed, multiplying the identity for $n = k$ by $X + Y$ and subtracting from it the product of the identity for $n = k - 1$ by XY, one obtains the identity for $n = k + 1$.

Therefore, on the one hand, we have

$$F_n\left(\phi, \frac{-1}{\phi}\right) = \sum_{i=0}^{\lfloor n/2 \rfloor} (-1)^i \binom{n-i}{i} \phi^i \left(\frac{-1}{\phi}\right)^i = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-i}{i}.$$

On the other hand, from the identity $X^{n+1} - Y^{n+1} = (X - Y)\sum_{i=0}^{n} X^i Y^{n-i}$, we obtain

$$F_n\left(\phi, \frac{-1}{\phi}\right) = \sum_{i=0}^{n} \phi^i \left(\frac{-1}{\phi}\right)^{n-i} = \frac{\phi^{n+1} - (-1/\phi)^{n+1}}{\phi + 1/\phi} = \frac{\phi^{n+1} - (-1/\phi)^{n+1}}{\sqrt{5}}.$$

Since $\sum_{i \geq 0} \binom{n-i}{i}$ satisfies the same recursion as the Fibonacci sequence and starts with F_2, F_3, it follows by induction that

$$\sum_{i \geq 0} \binom{n-i}{i} = F_{n+1} \quad \text{and one obtains} \quad F_{n+1} = \frac{\phi^{n+1} - (-1/\phi)^{n+1}}{\phi + 1/\phi},$$

which is Binet's formula.

This note was submitted in the beginning of 2001 and when it was accepted in October 2003, attention was drawn to two very enjoyable articles [**1, 2**] that appeared in the June 2003 issue. The authors studied a general Fibonacci-type of two-term linear recurrence:

$$g_{n+1} = ag_n + bg_{n-1},$$

where a, b are any (even complex!) constants. If we start with $g_0 = 1 = g_1$, then the analog of the formula

$$f_{n+1} = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-i}{i} \quad \text{is} \quad g_{n+1} = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-i}{i} b^i a^{n-2i},$$

as can be proved by induction. The corresponding Binet identity can be derived from the same polynomial identity above as follows. Consider the numbers defined by

$$\lambda + \mu = a, \quad \lambda\mu = -b.$$

$$\begin{aligned} F_n(\lambda, \mu) &= \sum_{i=0}^{\lfloor n/2 \rfloor} (-1)^i \binom{n-i}{i} (\lambda\mu)^i (\lambda + \mu)^{n-2i} \\ &= \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n-i}{i} b^i a^{n-2i} = g_{n+1}. \end{aligned}$$

Therefore,

$$g_{n+1} = \sum_{i=0}^{n} \lambda^i \mu^{n-i} = \frac{\lambda^{n+1} - \mu^{n+1}}{\lambda - \mu}.$$

This is Binet's formula for these general Fibonacci sequences.

It is fun to exploit the polynomial identity to derive some interesting identities involving binomial coefficients but the author would welcome a more natural motivation explaining the polynomial identity. Incidentally, one referee points out that Binet's formula appeared in De Morgan's notebooks before Binet was born.

Acknowledgment. The author is indebted to the referees for bringing to his notice the two beautiful articles cited.

REFERENCES

1. D. Kalman and R. Mena, The Fibonacci numbers—Exposed, this MAGAZINE **76**:3 (2003), 167–181.
2. A. T. Benjamin and J. J. Quinn, The Fibonacci numbers—Exposed more discretely, this MAGAZINE **76**:3 (2003), 182–192.

An Ideal Functional Equation with a Ring

ZORAN ŠUNIK
Department of Mathematics
Texas A&M University
College Station, TX 77843-3368
sunik@math.tamu.edu

There are many examples in mathematics and other sciences in which a single equation (or a relatively small system of equations) is capable of capturing the essence and complexity of an entire field. For example, the equation defining the zeroes of the Riemann zeta function plays a central role in analytic number theory and related fields. Even a partial understanding of the solutions to this equation would provide keys to the answers to many far-reaching questions that are currently in the research focus of the mathematical community. In a somewhat opposite direction, several outstanding results from the theory of elliptic curves and modular forms, crowning the efforts of several generations of mathematicians, were needed to tackle the single equation that is the subject of Fermat's Last Theorem. While the equation we present in this note does not live up to the high standards set by these two examples, it still has the same flavor in the sense that it touches on some very subtle questions and provides unexpected connections.

The ideal functional equation we refer to in the title and solve in the sequel is given by

$$f(xz-y)f(x)f(y)+3f(0)=1+2f(0)f(0)+f(x)f(y). \tag{1}$$

As it is, the equation looks far from ideal and its ring cannot be detected. In addition, it is not clear what it means to solve the given equation, so let us provide some context.

For example, we may solve the ideal equation over $\mathbb{R}$, which means that we need to find all functions $f:\mathbb{R}\to\mathbb{R}$ such that (1) is satisfied for all x, y, and z in $\mathbb{R}$.

Note that the ideal equation is equivalent to the following system of equations:

$$f(xz-y)f(x)f(y)=f(x)f(y), \tag{2}$$

$$f(0)=1. \tag{3}$$

Indeed, it is obvious that (2) and (3) together imply the ideal equation. On the other hand, setting $x=y=z=0$ in the ideal equation yields $(f(0)-1)^3=0$, which implies $f(0)=1$. Once we know that $f(0)=1$ the equation (2) easily follows from the ideal equation.

Now we try to solve the system. Let f be a function that satisfies the equations (2) and (3).

Setting $z=1$ and $y=0$ in (2) yields $f(x)f(x)=f(x)$. Thus, for all x, $f(x)=1$ or $f(x)=0$. Let $S=\{x \mid f(x)=1\}$. The function f is therefore given by

$$f(x)=\begin{cases}1, & \text{if } x\in S\\ 0, & \text{if } x\notin S,\end{cases} \tag{4}$$

that is, f is the characteristic function of the set S. However, not any set S would do. First, we know that $f(0)=1$, which means 0 must be in S. Further, set $z=1$ in (2). This gives

$$f(x-y)f(x)f(y)=f(x)f(y),$$

which implies that if both x and y are in S then so must be $x-y$. Now set $y=0$ in (2). This gives

$$f(xz)f(x)=f(x)$$

and therefore if x is in S then so must be xz, for any z. Thus, S has the following three properties:

$$\begin{aligned} &0\in S, \\ x,y\in S \quad &\text{implies} \quad x-y\in S, \\ x\in S \quad &\text{implies} \quad xz\in S, \quad \text{for all } z. \end{aligned} \tag{5}$$

It is easy to see that the only two subsets S of $\mathbb{R}$ that satisfy the three conditions in (5) are $\mathbb{R}$ itself and $\{0\}$. The characteristic functions, given by (4), of either $S=\mathbb{R}$ or $S=\{0\}$ satisfy the system and therefore also our ideal equation (1).

As we see, the solutions of our ideal equation over $\mathbb{R}$ are not particularly exciting. In order to get spicier solutions we consider the ideal equation over the integers $\mathbb{Z}$. To be precise, this means that we need to find all functions $f:\mathbb{Z}\to\mathbb{Z}$ that satisfy the equation (1) for all x, y, and z in $\mathbb{Z}$. For example, in addition to the characteristic functions of $\mathbb{Z}$ and $\{0\}$, the characteristic function of the set of even numbers

$$f(x) = \begin{cases} 1, & \text{if } x \text{ is even} \\ 0, & \text{if } x \text{ is odd} \end{cases}$$

satisfies the ideal equation. Indeed, if at least one of x or y is not even then both sides of (1) are equal to 3. If both x and y are even then both sides of (1) are equal to 4. Similarly, we can check that the characteristic function of any set $S = n\mathbb{Z} = \{nm \mid m \in \mathbb{Z}\}$ that consists of all multiples of a fixed integer n satisfies the ideal equation. In fact, by following the exact same steps as we did above for $\mathbb{R}$, we conclude that any solution to the ideal equation over $\mathbb{Z}$ must be the characteristic function of a subset S that satisfies the three conditions in (5). A subset of $\mathbb{Z}$ satisfies these conditions exactly when it consists of all multiples of a fixed integer n, that is, it is equal to $n\mathbb{Z}$, for some integer n.

The discussion so far might “ring” a bell. The conditions in (5) that describe the sets S providing the characteristic functions (4) that satisfy our “ideal” equation (1) are actually the conditions that define the notion of an *ideal* in a ring (for definitions of the terms *ring* and *ideal* see, for example, [**1**, pp. 194, 216]). It is easy to check that the relevant parts of our discussion remain valid for arbitrary integral domains [**1**, p. 200]. Thus, we make the following conclusion.

Conclusion A function $f : D \to D$ is a solution to the ideal equation (1) over an integral domain D exactly when it is the characteristic function, given by (4), of an ideal S of D.

The ideal structure in integral domains is a very important and difficult question that lies in the heart of several mathematical areas such as commutative algebra and algebraic geometry. It is quite interesting that we could capture all ideals in a single functional equation.

A few further modestly illuminating comments are in order.

We now see why our ideal equation does not have particularly exciting solutions over $\mathbb{R}$. The reason is that $\mathbb{R}$ is a field [**1**, p. 204] and every field has only two ideals, namely, the field itself and the zero ideal $\{0\}$.

If, instead of the ideal equation, we consider the more aesthetically appealing equation (2) alone, we note that its only solutions over an integral domain D are the characteristic functions of the ideals of D and the constant zero function. The fact that the zero function is a solution to (2) could be thought of as a nuisance, since this function is the characteristic function of the empty set, which is not an ideal of D. This is precisely why our ideal equation (1) had to be slightly more messy and include the part that served as a filter for the zero function.

Next, we note that we can use a similar functional equation to describe, say, the subrings [**1**, p. 196] of an integral domain D. One such equation is

$$f(x-y)f(xy)f(x) + 3f(0) = 1 + 2f(0)f(0) + f(x)f(y).$$

Once again, the messy part of the equation just serves to filter out the zero function and the interesting part is the equation

$$f(x-y)f(xy)f(x) = f(x)f(y),$$

which is satisfied by the characteristic functions of the subrings of the integral domain D as well as by the zero function.

If we do not insist on having a single equation describing the type of subsets we are interested in, and quite often there is no reason to do so, we can easily push our idea(l)s further. For example, the only solutions to the system of equations

$$f(0) = 1,\ f(1) = 0,$$
$$f(x-y)f(x)f(y) = f(x)f(y),$$
$$1 - f(xy) = \big(1 - f(x)\big)\big(1 - f(y)\big),$$

over an integral domain D are the characteristic functions of the prime ideals of D [**1**, p. 200]. We recall that a prime ideal S in a ring R is an ideal, different from R, that satisfies the condition

$$xy \in S \quad \text{implies} \quad x \in S \text{ or } y \in S.$$

In case, for aesthetic or some deeper reasons, we do insist on a single functional equation we may try the following. Assume that D is an integral domain for which there exists a polynomial $p(x_1, x_2)$ with coefficients in D such that

$$p(x_1, x_2) = 0 \quad \text{if and only if} \quad x_1 = x_2 = 0. \tag{6}$$

Then

$$p\big(x_1, p(x_2, x_3)\big) = 0 \quad \text{if and only if} \quad x_1 = x_2 = x_3 = 0,$$
$$p\Big(x_1, p\big(x_2, p(x_3, x_4)\big)\Big) = 0 \quad \text{if and only if} \quad x_1 = x_2 = x_3 = x_4 = 0,$$

and so on. Thus we can rewrite any system of n equations over D as a single equation over D. For example, the system

$$f_1 = 0, \quad f_2 = 0, \quad f_3 = 0,$$

could be rewritten as the single equation

$$p\big(f_1, p(f_2, f_3)\big) = 0.$$

A polynomial $p(x_1, x_2)$ with the property (6) exists in many cases. A precise description of all such cases would lead us considerably deeper in the subject than we are willing to go at this moment. As an easy example, we note that such a polynomial over $\mathbb{Z}$ is $p(x_1, x_2) = x_1^2 + x_2^2$. Thus, we may write a single functional equation that describes all prime ideals of $\mathbb{Z}$. Since the prime ideals of $\mathbb{Z}$ are the zero ideal $\{0\}$ and the ideals of the form $p\mathbb{Z} = \{pm \mid m \in \mathbb{Z}\}$, for p a prime number, such a functional equation, rather implicitly, describes the prime numbers. With the last observation, we just made a quick peek into number theory.

Finally, we conclude our little interplay of ring theory, functional equations, and logic with an invitation to try to find functional equations leading to other types of subsets in integral domains and, more generally, to various substructures in other algebraic settings. A particularly easy example is the case of Boolean algebras [**1**, p. 511]. If B is a Boolean algebra, the only functions $f : B \to B$ that satisfy the equation

$$f(0) \wedge \big[f(x) \Rightarrow f(\bar{x})\big] \wedge \big[(f(x) \wedge f(y)) \Rightarrow f(x \wedge y)\big] = 1, \tag{7}$$

for all x and y in B, are exactly the characteristic functions of the Boolean subalgebras of B (an expression of the form $A \Rightarrow B$ used in (7) is just a shorthand for $\bar{A} \vee B$).

REFERENCES

1. Aigli Papantonopoulou, *Algebra, Pure and Applied*, Prentice Hall, New Jersey, 2002.

Wieferich Primes and Period Lengths for the Expansions of Fractions

GENE GARZA
JEFF YOUNG
University of Montevallo
Montevallo, Al 35115
genesr@prodragon.com

It is well known that some decimal expansions terminate, while others repeat, at least eventually, in patterns, which may be short or lengthy (we shall call this repeating pattern the *period* of the expansion). Here we will extend some known results while exploring expansions of fractions in any base. Our goal will be to find a formula for the length of the period of such expansions. The interested reader is referred to the recent award-winning article by Jones and Pearce, who show how to display such decimal expansions graphically [**3**].

We will consider both the expansions of (the reciprocals of) primes and of composites. It would seem that the easier part of this problem would be that of primes. However, there are difficulties/anomalies among primes that make it hard to find a formula that works in all cases. The most interesting such case is that of Wieferich primes, whose reciprocals are characterized by expansions whose periods are the same length as the periods of their squares. For example, the length of the period of $1/1093$ is 1092 which is the same as that of $1/1093^2$. This, as we shall see, is not normally the case. For someone seeking a simple formula, this is bad news. However, as our table at the end shows, Wieferich primes are quite rare.

Preliminaries Let's review what is meant by the expansion of a fraction and, in particular, the decimal expansion of a fraction. A few examples should suffice. In what follows, a line over digits in a decimal expansion (or expansion in any base b) will denote that those digits repeat infinitely often in that expansion.

$$\begin{aligned}
1/3 &= 0.\overline{3} && \text{(period 1, base 10)}\\
1/3 &= 0.\overline{01} && \text{(period 2, base 2)}\\
1/9 &= 0.\overline{1} && \text{(period 1, base 10)}\\
1/9 &= 0.\overline{000111} && \text{(period 6, base 2)}\\
1/27 &= 0.\overline{037} && \text{(period 3, base 10)}\\
1/27 &= 0.\overline{000010010111101101} && \text{(period 18, base 2)}
\end{aligned}$$

We say that $0.\overline{3} = 0.333\cdots$ is the expansion for $1/3$ in base 10 (decimal), that $0.\overline{01} = 0.010101\cdots$ is the expansion for $1/3$ in base 2, etc. The expansions in bases other than 10 can be obtained by long division after converting to the new base.

PROPOSITION. *The* period *of the expansion of* $1/x$*, in base b, is the smallest number, say p, for which* $b^p \equiv 1 \pmod{x}$.

That is, the period is the smallest number p such that $x \mid b^p - 1$. (This is basically Th. 4, section 15 from Dudley's book [**2**].)

DEFINITION 1. *By the* period of a number, *x, in base b, we shall mean the period of the expansion of* $1/x$ *in base b.*

When considering expansions, it will be our intention to concentrate on just the expansions of reciprocals of integers. This is sufficient since the length of the period of a fraction depends only on the denominator as long as the numerator is relatively prime to the denominator. To see this, consider the following in base ten:

$$\begin{aligned}
1/7 &= 0.\overline{142857} \\
10/7 &= 1 + 3/7 &&= 1.\overline{428571} \\
100/7 &= 14 + 2/7 &&= 14.\overline{285714} \\
1000/7 &= 142 + 6/7 &&= 142.\overline{857142} \\
10000/7 &= 1428 + 4/7 &&= 1428.\overline{571428} \\
100000/7 &= 14285 + 5/7 &&= 14285.\overline{714285}
\end{aligned}$$

Clearly then, the length of the expansion for any proper fraction with denominator 7 is 6. Different numerators simply serve to change the starting digit of the period. For other bases b, we need only note that $b = 10_b$; that is, in base b, b is 10. Thus, for any given base, a reduced proper fraction with denominator x will have a period of the same length as $1/x$.

Now consider the decimal expansions for 3, 6, 15, and 30:

$$\begin{aligned}
1/3 &= 0.\overline{3} \\
1/6 &= 0.1\overline{6} \\
1/15 &= 0.0\overline{6} \\
1/30 &= 0.0\overline{3}
\end{aligned}$$

These expansions suggest that factors of the base in the denominators do not affect the length of the period, but only delay its beginning. This is easily seen in the following example:

$$\begin{aligned}
1/7 &= 0.\overline{142857} \\
1/35 &= 2/(10 \cdot 7) = .0\overline{285741} \\
1/14 &= 5/(10 \cdot 7) = 0.0\overline{714285}
\end{aligned}$$

Thus, when looking for the length of the period for an expansion it is enough to factor out all numbers from the denominator that divide the base, and determine the length of the period for the remaining number.

Periods of composites It seems natural to ask about the periods of composites whose factors may or may not be repeated and whose factors include none of the factors of the base. Some of these questions have been answered, and it is our purpose to consider these questions and to provide some additional answers.

First of all, it is well known that the expansion of a composite whose prime factors are not repeated and are not factors of the base has a period length that is just the lcm (least common multiple) of the periods of the individual factors [**2**]. For example, the period of 77 in base 10 is 6, since the period of 7 is 6, the period of 11 is 2, and $\operatorname{lcm}(6, 2) = 6$. Similarly, the period of $341 = 11 \cdot 31$ is 30, since the period of 11 is 2 and the period of 31 is 15.

The "lcm rule" makes such problems quite manageable. It remains to consider powers of single primes. A few examples would again be useful. In base 10,

$$\begin{aligned}
1/7 &= 0.\overline{142857} && \text{(period 6)} \\
1/7^2 &= 0.\overline{020408163265306122448979591836734693877551} && \text{(period 42)} \\
1/7^3 &= 0.\overline{0029155\cdots} && \text{(period 294).}
\end{aligned}$$

In base 2,

$$\begin{array}{ll} 1/7 = 0.\overline{001} & \text{(period 3)} \\ 1/7^2 = 0.\overline{000001010011100101111} & \text{(period 21)} \\ 1/7^3 = 0.\overline{000000001\cdots} & \text{(period 147).} \end{array}$$

Careful observation leads one to conjecture that the period for, say x^n, when x is prime, is just the period of x multiplied by x^{n-1}. The unfortunate difficulty with attempting to prove this *power rule* conjecture is that it is not true! Counterexamples are abundant; just look at $1/3$, $1/9$, and 127 in base 10. The periods for the expansions of these numbers are 1, 1, and 3, respectively.

However, there is something special about 9 in base 10, which will eventually lead us to refine our conjecture. Actually, for any base b there is something special about the expansion of $1/(b-1)$. One can see this by considering the following geometric series in base b:

$$\frac{1}{b-1} = \frac{1}{b} + \frac{1}{b^2} + \frac{1}{b^3} + \cdots$$

In "decimal point" notation for base b, the expansion for $b-1$ is nothing more than $.111\cdots$. Because of this, any factor of $b-1$ has a period of length 1 in base b.

To illustrate, let $b-1$ be the product of, say, x and y (which are both less than b, and therefore are just "digits" in base b), and consider the expansion in base b of $1/x$. It is not too hard to see that it is nothing more than $.yyy\cdots$. For example in base 11, $1/2 = .555\cdots$ and $1/5 = .222\cdots$. This may be verified by observing that $1/2$ is nothing more than $1/2 = 5/11 + 5/11^2 + 5/11^3 \cdots$. Likewise, $1/5 = 2/11 + 2/11^2 + 211^3 \cdots$.

In base $b = 10$ the only factors of $b-1$ are 3 and 9. Here, we note that $10^1 \equiv 1 \pmod 3$ and that $10^1 \equiv 1 \pmod{3^2}$.

Period one primes For a particular base b, factors of $b-1$, which we call *period one primes*, provide counterexamples to our conjectured power rule formula. However, in any given base, period one primes will obviously be scarce so we simply eliminate all period one primes from consideration. In base 10, 3 is easily verified as the only period one prime. Of course, base 2 has no period one primes.

If we eliminate all period one primes and reconsider our conjecture, we are once again doomed—but for a different reason. As we shall see shortly, there are certain exceptions that occur, perhaps in all bases. So, what can we say with confidence? Well, it is certainly true, as we shall prove for any prime x and any base b, that $b^{px^{n-1}} \equiv 1 \pmod{x^n}$, where p is the period of x in base b.

Of course, this does not mean that px^{n-1} is actually the length of the period for the expansion of $1/x^n$. It is well known [2] that the period of x^n must divide $b^{px^{n-1}} - 1$, but we do not know that px^{n-1} is the smallest such number. What might happen in this case? In our earlier efforts to prove the power rule, the difficulty always occurred at the same point. It seemed unlikely at first that some x^2 might divide $b^p - 1$, where p is the period of x, since this would imply that the period of x^2 is the same as the period of x. This brings us to the *Wieferich primes*.

Wieferich primes A *Wieferich prime in base b* is a prime number, x that satisfies the congruence

$$b^{x-1} \equiv 1 \pmod{x^2}.$$

(In some discussions, the base b is limited to be 2.) Are they common? Do they exist in all bases? The answers to these questions are not all known. However, it is known [**5**] that Wieferich primes exist for many different bases, and we offer a table of Wieferich primes at the end of this Note.

In base 2, for example, 1093 and 3511 are Wieferich primes. This means that not only is $2^{1092} \equiv 1 \pmod{1093}$ and $2^{3510} \equiv 1 \pmod{3511}$ (which follows by Fermat's Little Theorm [**2**]), but also that $2^{1092} \equiv 1 \pmod{1093^2}$ and $2^{3510} \equiv 1 \pmod{3511^2}$. Crandall, Dilcher, and Pomerance [**1**] showed in 1997 that the only base-2 Wieferich primes below $4 \cdot 10^{12}$ are 1093 and 3511.

Actually, we will characterize Wieferich primes slightly differently. Of course, since $x - 1$ must be divisible by the period p, we change this definition to primes characterized by $b^p \equiv 1 \pmod{x^2}$. (In light of the upcoming corollary with $mq = x - 1$ and $n = 2$, we see that the new definition is equivalent to the previous one.) We note that if $b^p \equiv 1 \pmod{x^n}$ where n is anything higher than 2, then it is also true that $b^p \equiv 1 \pmod{x^2}$. Thus, for our purposes, if $b^p \equiv 1 \pmod{x^3}$ then x is a Wieferich prime for base b. We shall also refer to Wieferich primes as "primes with square periods" to emphasize the exceptional cases where the periods of the expansions for $1/x$ are the same as the periods of the expansions for their squares, $1/x^2$.

There are, of course, period one numbers with not only square periods, but cube periods and even higher. To see this, consider $9^1 \equiv 1 \pmod 2$, $9^1 \equiv 1 \pmod{2^2}$, $9^1 \equiv 1 \pmod{2^3}$. Here the period for each of 2, 2^2 and 2^3 is 1 in base 9. Two better examples might be $3^{10} \equiv 1 \pmod{11^2}$ and $7^4 \equiv 1 \pmod{5^2}$, where $3^{10} \equiv 1 \pmod{11}$ and $7^4 \equiv 1 \pmod 5$. It is thus apparent that if one is to compute, by way of some formula, the period for an expansion in any base, then those rare numbers with square periods must be considered and discounted. Indeed, we shall derive such a formula for the length of the period of a number whenever period one numbers and numbers with square periods are discarded. We will call this formula by the obvious name, the *power rule*.

The power rule Before stating our main theorem we need the following lemmas and corollaries:

LEMMA 1. *Suppose* $a_i \equiv 1 \pmod x$ *for each* a_i, $i = 1, \ldots, m$, *where* $m > 0$. *Then* $\sum a_i \equiv 0 \pmod x$ *if and only if* $m \equiv 0 \pmod x$.

The proof is left as an exercise for the reader.

COROLLARY. *If* $b^{mq} \equiv 1 \pmod{x^n}$ *for* $n \geq 1$ *where* q *is a multiple of the period of* x, *but* m *is not a multiple of* x, *then* $b^q \equiv 1 \pmod{x^n}$.

Proof. To see this we will rewrite $(b^{mq} - 1)$ as $(b^q)^m - 1$ and write q as dp where d is an integer. Then we factor $b^{mq} - 1 = b^{pmd} - 1$ as $(b^{pd} - 1)(b^{pd(m-1)} + b^{pd(m-2)} + \cdots + 1) = (b^q - 1)(b^{pd(m-1)} + b^{pd(m-2)} + \cdots + 1)$. Applying Lemma 1 to the second factor, which has m terms, each of which is congruent to $1 \pmod x$ (since p is the period of x), we see that x^n must divide $(b^q - 1)$ so that $b^q \equiv 1 \pmod{x^n}$. ∎

LEMMA 2. *If* x *is an odd prime,* $k > 1$, *and* $b^{px^{(k-1)}} \equiv 1 \pmod{x^{k+1}}$ *where* p *is the period of* x *in base* b, *then* $b^{px^{(k-2)}} \equiv 1 \pmod{x^k}$. (Note that x is selected to be odd, since 2 is period one for all odd bases. Otherwise, $7^2 \equiv 1 \pmod{2^4}$ while 7^1 is not 1 $\pmod{2^3}$ would be an obvious exception.)

Proof. Since p is the period of x, we have $b^p \equiv 1 \pmod x$ and we can write b^p as $(1 + nx)$ for some integer n. By the Binomial Theorem,

(1) $(b^p)^{x^{(k-1)}} \equiv 1 + nx^k \pmod{x^{k+1}}$ and

(2) $(b^p)^{x^{(k-2)}} \equiv 1 + nx^{k-1} \pmod{x^k}$.

But, by our hypothesis, $b^{px^{(k-1)}} \equiv 1 \pmod{x^{k+1}}$. This, along with (1) implies that $x \mid n$. The conclusion that $b^{px^{(k-2)}} \equiv 1 \pmod{x^k}$ then follows from (2) and the proof is complete. ■

The idea behind Lemma 2 is that under certain conditions factors of x can be cancelled from congruences. Now we are prepared to state and prove our main result:

POWER RULE THEOREM. *If x is an odd prime, $N = x^n$, $n > 1$, and x is not a period one prime nor a Wieferich prime for base b, that is, not a prime with a square period, then the period of N is px^{n-1} where p is the period of x.*

Proof. We need to show two things:

I. $x^n \mid b^{px^{(n-1)}} - 1$.

II. If $x^n \mid b^Q - 1$, then $Q \geq px^{n-1}$.

I. This follows immediately from the binomial theorem since $b^p = 1 \pmod{x}$.

II. We must show for $n \geq 2$ that if $x^n \mid b^Q - 1$, then $Q \geq px^{n-1}$ for since we already know that it is true for $n = 1$. Assume not, that is, assume $Q < px^{n-1}$.

Once again, we know that Q must be a multiple of p, the period of x, since $x^n \mid b^Q - 1$ implies $x \mid b^Q - 1$. So let $Q = mp$. There are two cases to consider.

First, let's consider the case where m is a multiple of x. We write $m = rx^t$ where $1 \leq t < n - 1$ and r is not a multiple of x, so that $Q = rx^t p$. Here we have $x^n \mid b^{rx^t p} - 1$. By the Corollary, we can cancel the r so that $x^n \mid b^{px^t} - 1$. By Lemma 2, we can cancel one x from both sides of the expression to obtain $x^{n-1} \mid b^{px^{t-1}} - 1$. This we may repeat until we have $x^{n-t} \mid b^p - 1$ since under our assumptions $n - t \geq 2$. But this means that x is a Wieferich prime contrary to our hypotheses, so we conclude that m is not a multiple of x.

Second and finally, we consider what happens when m is not a multiple of x. In this case we have $x^n \mid b^{mp} - 1$. Once again using the Corollary, we can cancel m so that $x^n \mid b^p - 1$. Since $n \geq 2$, we conclude that x must be a Wieferich prime, which once again violates our hypothesis. This completes the proof of the Power Rule Theorem. ■

Conclusion Together with the lcm rule, the power rule provides a formula for the period of the expansion for the reciprocal of any composite—as long as no factors of the composite are to be excluded such as Wieferich primes or period one numbers. This formula may be evaluated easily as long as the periods for the individual prime factors are known. Predicting the periods for an arbitrary prime is, however, still elusive. A table of Wieferich primes is provided to demonstrate their scarcity in bases up to 25 for numbers up to 2^{18}. (Period one numbers that qualify as Wieferich primes, such as 2 in base 9, have been removed from the table.) This table contains two particularly interesting entries: $18^6 \equiv 1 \pmod{7^2}$ and $19^6 \equiv 1 \pmod{7^2}$. The interesting part is that the following congruences are also valid: $18^6 \equiv 1 \pmod{7^3}$ and $19^6 \equiv 1 \pmod{7^3}$. This shows that there exist non period-one Wieferich primes with cube periods—in this case, for bases 18 and 19. For other bases this is still an open question [**5**]. Not quite so obvious from the table is the fact that $18^3 \equiv 1 \pmod{7^2}$ and $18^3 \equiv 1 \pmod{7^3}$. This answers, negatively, the question [**5**] whether Wieferich primes, x, must have periods of maximal length, that is, $x - 1$. Another such example is: $3^{10} \equiv 1 \pmod{11^2}$ and $3^5 \equiv 1 \pmod{11^2}$.

TABLE 1: Table of Wieferich primes up to 2^{18} for bases up to 25.

base	Wieferich primes	base	Wieferich primes
2	1093, 3511	14	29, 353
3	11	15	29131
4	1093, 3511	16	1093, 3511
5	20771, 40487	17	3, 46021, 48947
6	66161	18	5, 7, 37, 331, 33923
7	5	19	7, 13, 43, 137
8	3, 1093, 3511	20	281, 46457
9	11	21	None
10	487	22	13, 673
11	71	23	13
12	2693, 123653	24	5, 25633
13	863	25	20771, 40487

Open Questions To repeat, it is known that Wieferich primes satisfying the congruence, $b^p \equiv 1 \pmod{x^2}$, exist and are rare, but it is not known if any Wieferich primes, other than period one primes, satisfy the congruence $b^p \equiv 1 \pmod{x^n}$, $n > 2$ except for bases 18 and 19. Also, it is not known if there are Wieferich primes for each base; or even if the set of such primes is infinite (discounting period one primes, once again).

CONJECTURE. Wieferich primes exist for all bases and, furthermore, the following relationship holds in any base, b, for infinitely many primes, x, and for any value of n: $b^{x-1} \equiv 1 \pmod{x^n}$.

Acknowledgment. The authors would like to express their appreciation to the reviewers for many suggestions and improvements.

REFERENCES

1. Richard Crandall, Karl Dilcher, and Carl Pomerance, A search for Wieferich and Wilson primes, *Math. Comp.* **66** (1997), 433–449.
2. Underwood Dudley, *Elementary Number Theory*, W. H. Freeman and Company, San Francisco, 1969.
3. Rafe Jones and Jan Pearce, A postmodern view of fractions and the reciprocals of Fermat primes, this MAGAZINE **73**:2 (2000), 83–96.
4. William Levitt, Repeating decimals, *College Math. J.* **15** (1984), 299–308.
5. P. Ribenboim, *The Book of Prime Number Records*, 2nd ed., Springer, New York, 1989.

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

Proposals

To be considered for publication, solutions should be received by March 1, 2005.

1701. *Proposed by Murray S. Klamkin, University of Alberta, Edmunton, AB.*

Prove that for all positive real numbers a, b, c, d,

$$a^4b + b^4c + c^4d + d^4a \geq abcd(a + b + c + d).$$

1702. *Proposed by Roy Barbara, American University of Beirut, Beirut, Lebanon.*

Let R be the circumradius of nondegenerate triangle ABC. For point P on or inside of triangle ABC, let $S(P) = |PA| + |PB| + |PC|$. Find the maximum value of k and the minimum value of K such that $kR \leq S(P) \leq KR$ for all acute triangles ABC.

1703. *Proposed by Shahin Amrahov, ARI College, Ankara, Turkey.*

Let $p(x) = ax^3 + bx^2 + cx + d$ where a, b, c, d are integers with $a \equiv c \equiv 2 \pmod 3$ and $b \equiv 0 \pmod 3$. Prove that for any positive integer n there is an integer k such that $p(k)$ is a multiple of 3^n.

1704. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let F be the Fermat-Torricelli point for triangle ABC. Let the cevian from B through F meet $\overline{AC}$ in B^* and the cevian from C through F meet $\overline{AB}$ in C^*. Prove that if $BB^* = CC^*$, then triangle ABC is isosceles. (The Fermat-Torricelli point F of triangle ABC is the point for which the sum of the distances from the vertices is minimum.)

We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals must, in general, be accompanied by solutions and by any bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution.

Solutions should be written in a style appropriate for this MAGAZINE. Each solution should begin on a separate sheet.

Solutions and new proposals should be mailed to Elgin Johnston, Problems Editor, Department of Mathematics, Iowa State University, Ames IA 50011, or mailed electronically (ideally as a LaTeX file) to ehjohnst@iastate.edu. All communications should include the readers name, full address, and an e-mail address and/or FAX number.

1705. *Proposed by Michel Bataille, Rouen, France.*

Let n be a positive integer. Find the minimum value of

$$\frac{(a-b)^{2n+1}+(b-c)^{2n+1}+(c-a)^{2n+1}}{(a-b)(b-c)(c-a)},$$

for distinct real numbers a, b, c with $bc + ca \geq 1 + ab + c^2$.

Quickies

Answers to the Quickies are on page 326.

Q943. *Proposed by Michael Botsko, Saint Vincent College, Latrobe, PA.*

Let (X, d) be a compact metric space and let $f : X \longrightarrow X$ be an onto function. Suppose that $d(f(x), f(y)) > d(x, y)$ whenever $x \neq y$. Prove that f has a unique fixed point.

Q944. *Proposed by Michel Bataille, Rouen, France.*

Show that for positive integer n,

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}\binom{2n-k}{n} = 1.$$

Solutions

In Crowded Circles **October 2003**

1677. *Proposed by Mowaffaq Hajja, Yarmouk University, Irbid, Jordan.*

Let triangle ABC be inscribed in circle $\mathcal{C}$. Let A^* be the point on $\mathcal{C}$ that bisects the arc BC that contains A. Let B^* and C^* be defined in a similar way.

(a) Prove that $\overline{A^*B^*}$ is parallel to $\overline{C^*C}$, $\overline{B^*C^*}$ is parallel to $\overline{A^*A}$, and $\overline{C^*A^*}$ is parallel to $\overline{B^*B}$.

(b) The points A, B, C, A^*, B^*, C^* partition $\mathcal{C}$ into six arcs. The length of the shortest of these can be taken as a measure of the "crowdedness" of the six points. How should A, B, C be chosen so the points A, B, C, A^*, B^*, C^* are least crowded?

Solution by Mike Hitchman, Albertson College, Caldwell, ID.

We assume that $\triangle ABC$ is scalene. Without loss of generality we may also assume that $\mathcal{C}$ is the unit circle, that $A = (1, 0)$, and that $AB < BC$ and $AB < CA$. By possibly reflecting about the x-axis and about the perpendicular bisector of AB, we may assume that $B = (\cos\beta, \sin\beta) = \operatorname{cis}\beta$ with $0 \leq \beta < \frac{2\pi}{3}$ and that $C = \operatorname{cis}\gamma$ where $2\beta < \gamma < \frac{\beta}{2} + \pi$. We then have

$$A^* = \operatorname{cis}\left(\pi + \frac{\beta+\gamma}{2}\right), \qquad B^* = \operatorname{cis}\frac{\gamma}{2}, \quad \text{and} \quad C^* = \operatorname{cis}\left(\pi + \frac{\beta}{2}\right).$$

In counterclockwise order around C, the six points are in order A, B, B^*, C, C^*, A^*. For the remainder of the solution, $\text{arc}(PQ)$ will denote the arc on the circle that runs counterclockwise from P to Q.

(a) The midpoint of $\text{arc}(CC^*)$ is

$$\text{cis}\left(\frac{\pi+\gamma+\beta/2}{2}\right).$$

This is also the midpoint of $\text{arc}(A^*B^*)$. Because the midpoints of the arcs coincide, it follows that $\overline{A^*B^*}$ is parallel to $\overline{CC^*}$. A similar argument establishes the other two cases.

(b) We show that the six points are least crowded when $A = 1 = \text{cis}(0)$, $B = \text{cis}(2\pi/9)$, and $C = \text{cis}(8\pi/9)$. The measures of the six arcs are

$$\text{arc}(AB) = \beta, \quad \text{arc}(BB^*) = \frac{\gamma}{2} - \beta, \quad \text{arc}(B^*C) = \frac{\gamma}{2}, \quad \text{arc}(CC^*) = \pi + \frac{\beta}{2} - \gamma,$$

$$\text{arc}(C^*A^*) = \frac{\gamma}{2}, \quad \text{and} \quad \text{arc}(A^*A) = \pi - \frac{\gamma+\beta}{2}.$$

In addition, we have the following arc length relationships,

(1) $\text{arc}(AB) + \text{arc}(BB^*) + \text{arc}(B^*C) + \text{arc}(CC^*) + \text{arc}(C^*A^*) + \text{arc}(A^*A) = 2\pi$

(2) $\text{arc}(A^*A) = \text{arc}(BB^*) + \text{arc}(CC^*)$

(3) $\text{arc}(B^*C) = \text{arc}(C^*A^*) = \text{arc}(AB) + \text{arc}(BB^*)$.

Equations (2) and (3) ensure that $\text{arc}(A^*A)$, $\text{arc}(B^*C)$, and $\text{arc}(C^*A^*)$ all have measure greater than the measure of $\text{arc}(BB^*)$. In addition, by substituting from (2) and (3) into (1) we find

$$3\,\text{arc}(AB) + 4\,\text{arc}(BB^*) + 2\,\text{arc}(CC^*) = 2\pi.$$

It follows that the six points are least crowded when these three arcs all have the same measure, namely, $2\pi/9$. In particular, this is the case when $\text{arc}(AB) = 2\pi/9$, $\text{arc}(BC) = \text{arc}(BB^*) + \text{arc}(B^*C) = \text{arc}(AB) + 2\,\text{arc}(BB^*) = 2\pi/3$, and $\text{arc}(CA) = 10\pi/9$, so $\triangle ABC$ has angles of $\pi/9$, $\pi/3$, and $5\pi/9$.

Note. If $\triangle ABC$ is not scalene, then at least one of the point pairs X, X^* will coincide. Part (a) still holds provided the segment XX^* is not degenerate.

Also solved by Herb Bailey, Roy Barbara (Lebanon), Michel Bataille (France), Daniele Donini (Italy), Michael Golgenberg and Mark Kaplan, David Gove, Peter J. Gressis, Enkel Hysnelaj (Australia), Victor Y. Kutsenok, Li Zhou, and the proposer. There were six incorrect or incomplete submissions.

The Ladder and the Box **October 2003**

1678. *Proposed by Kent Holing, Statoil Research Center, Trondheim, Norway.*

Let $P[h, w, l]$ denote the following problem: *A box of height h and width w is placed on the floor so it is flush with a wall and the sides of length w are perpendicular to the wall. A ladder of length l leans against the wall and just touches the box along the upper edge. Let x be the length of the portion of the ladder that is between the wall and the point of contact with the box.*

(a) Characterize all ordered triples (h, w, l) of positive integers such that the problem $P[h, w, l]$ has at least one integral solution for x.

(b) Characterize all ordered triples (h, w, l) of positive integers such that the problem $P[h, w, l]$ has exactly one solution for x, and that solution is integral.

Solution by Knut Dale, Telemark University College, Bø, Norway.

If the ladder touches the floor, the wall, and the box at one point, and makes angle $\theta \in (0, \pi/2)$ with the floor, then

$$l = \frac{h}{\sin\theta} + \frac{w}{\cos\theta}.$$

As θ increases from 0 to $\pi/2$, the value of l decreases from ∞ to a minimum $l = l_0$ characterized by $l_0^{2/3} = h^{2/3} + w^{2/3}$ (when $\tan^3\theta = h/w$) then increases to ∞. If $l > l_0$, then there are two positions in which the ladder touches the box, and each leads to a solution for x. If $l = l_0$ then there is just one position in which the ladder touches the box and just one solution for x. If $l < l_0$ then there are no solutions.

(a) First assume that $l \geq l_0$. We show that $P[h, w, l]$ has at least one positive integer solution if and only if $(h, w, l) = (\alpha r, \beta s, (\alpha + \beta)t)$, where (r, s, t) is a primitive Pythagorean triple and α, β are positive integers.

Assume the problem has an integer solution x, and let a and b be, respectively, the distances from the top and the foot of the ladder to the box. By similar triangles

$$\frac{h}{a} = \frac{b}{w} = \frac{l - x}{x}.$$

If $h, w, l, x \in \mathbb{N}$, then $a, b \in \mathbb{Q}^+$. Because $a^2 = x^2 - w^2$ and $b^2 = (l - x)^2 - h^2$, it follows that $a, b \in \mathbb{N}$. Hence $(h, b, l - x)$ and (a, w, x) are both Pythagorean triples and are associated with similar triangles. Thus there are a primitive Pythagorean triple (r, s, t) and $\alpha, \beta \in \mathbb{N}$ with $(h, b, l - x) = \alpha(r, s, t)$ and $(a, w, x) = \beta(r, s, t)$. Then $(h, w, l) = (\alpha r, \beta s, (\alpha + \beta)t)$ and one solution to the problem $P[h, w, l]$ is the positive integer $x = \beta t$. Conversely, given such a representation for (h, w, l), there is at least one integer solution for x, namely $x = \beta t$.

(b) The problem has exactly one solution x if and only if $l = l_0$. We show that in this case, h, w and $l = l_0$ are all positive integers if and only if $(h, w, l) = (\gamma r^3, \gamma s^3, \gamma t^3)$ where (r, s, t) is a primitive Pythagorean triple and $\gamma \in \mathbb{N}$. For such (h, w, l), $x = \gamma s^2 t$ and is integral.

Suppose that $l^{2/3} = h^{2/3} + w^{2/3}$ for some $h, w, l \in \mathbb{N}$. Rewrite the equation in the form $h = (hl^2)^{1/3} - (hw^2)^{1/3}$. Set $(hl^2)^{1/3} = y + h/2$ and $(hw^2)^{1/3} = y - h/2$. Then $hl^2 - hw^2 = 3hy^2 + \frac{1}{4}h^3$, showing that $y^2 \in \mathbb{Q}$, and from $hl^2 + hw^2 = y(2y^2 + \frac{3}{2}h^2)$ we find $y \in \mathbb{Q}$. Hence, $hl^2 = p^3$ and, by symmetry, $wl^2 = q^3$ for some $p, q \in \mathbb{Q}$, and it then follows that $p, q \in \mathbb{N}$. Let $d = \mathrm{GCD}(p, q)$, so $p = dr$ and $q = ds$ with $(r, s) = 1$. We then find $hs^3 = wr^3$, so $h = \gamma r^3$ and $w = \gamma s^3$ for some $\gamma \in \mathbb{N}$. Finally, because $l^2 = \gamma^2(r^2 + s^2)^3$, we conclude that $l = \gamma t^2$ with $r^2 + s^2 = t^2$. This shows that (h, w, l) has the desired form. Conversely, if (h, w, l) has this form, then $l^{2/3} = h^{2/3} + w^{2/3}$ so there is exactly one solution and the solution, $x = \gamma s^2 t = w^{2/3} l^{1/3}$, is integral.

Also solved by Roy Barbara (Lebanon), Michel Bataille (France), John Christopher, Con Amore Problem Group (Denmark), Chip Curtis, Martin Levin, H. T. Tang, Li Zhou, and the proposer. There were four incomplete or incorrect submissions.

Fibonacci Power **October 2003**

1679. *Proposed by José Luis Díaz-Barrero and Juan José Egozcue, Universitat Politècnica de Catalunya, Barcelona, Spain.*

Let f_k denote the kth Fibonacci number, that is, $f_0 = 0$, $f_1 = 1$, and $f_{k+2} = f_{k+1} + f_k$, $k \geq 0$. Prove that for any positive integer n,

$$\sum_{k=1}^{n} \binom{n}{k} \log\left(f_k^{f_{2n}}\right) \leq (2^n - 1) \sum_{k=1}^{n} \binom{n}{k} \log\left(f_k^{f_k}\right).$$

Solution by G.R.A.20 Problems Group, Rome, Italy.

Using the identities

$$\sum_{k=1}^{n} \binom{n}{k} = 2^n - 1 \quad \text{and} \quad \sum_{k=1}^{n} \binom{n}{k} f_k = f_{2n},$$

and the fact that $-\ln x$ is convex,

$$\begin{aligned}\frac{1}{2^n - 1} \sum_{k=1}^{n} \binom{n}{k} \ln(f_k) &\leq \ln\left(\frac{1}{2^n - 1} \sum_{k=1}^{n} \binom{n}{k} f_k\right) \\ &= \ln\left(\frac{f_{2n}}{2^n - 1}\right) = -\ln\left(\frac{1}{f_{2n}} \sum_{k=1}^{n} \binom{n}{k} f_k \frac{1}{f_k}\right) \\ &\leq \frac{1}{f_{2n}} \sum_{k=1}^{n} \binom{n}{k} f_k \left(-\ln\left(\frac{1}{f_k}\right)\right) = \frac{1}{f_{2n}} \sum_{k=1}^{n} \binom{n}{k} f_k \ln(f_k).\end{aligned}$$

This is equivalent to the inequality in the problem statement.

Also solved by Michel Bataille (France), Minh Can, Daniele Donini, Ovidiu Furdui, Enkel Hysnelaj (Australia), Heinz-Jürgen Seiffert (Germany), Achilleas Sinefakapoulos, Ricardo M. Torrejón, Li Zhou, and the proposer. There were two incorrect submissions.

Symmetric Extremes **October 2003**

1680. *Proposed by H. A. Shah Ali, Tehran, Iran.*

For positive integers k and n, with $k \leq n$, define the kth elementary symmetric function $S_{k,n}$ of the n numbers $x_1, x_2, \ldots, x_n$ by

$$S_{k,n} = S_{k,n}(x_1, \ldots, x_n) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k}.$$

Let $c \in (0, 1]$ be given and assume that $x_i \geq 0$ for $1 \leq i \leq n$, and that $\sum_{k=1}^{n} S_{k,n} = c$. Find the minimum and maximum values of $S_{k,n}$ for each $1 \leq k \leq n$, and determine necessary and sufficient conditions for the extrema to occur.

Solution by Li Zhou, Polk Community College, Winter Haven, FL.

For $1 \leq k \leq n$, let $m_{k,n}$ and $M_{k,n}$ be, respectively, the minimum and maximum values of $S_{k,n}$. We first observe that $m_{1,1} = c = M_{1,1}$ with $x_1 = c$. For the remainder of the solution we assume that $n \geq 2$.

The key to our solution is the replacement of pairs (x_i, x_j), $i \neq j$, by $(y, 0)$ or (z, z) where $1 + y = (1 + x_i)(1 + x_j)$ and $1 + z = \sqrt{(1 + x_i)(1 + x_j)}$. Because

$$1 + \sum_{k=1}^{n} S_{k,n} = \prod_{i=1}^{n} (1 + x_i),$$

these replacements do not change the value of $\sum_{k=1}^{n} S_{k,n}$. Observe also that

$$y = x_i + x_j + x_i x_j \geq x_i + x_j,$$

with equality if and only if at least one of x_i, x_j is 0. Also, $2z + z^2 = x_i + x_j + x_i x_j$ and, by the arithmetic-geometric mean inequality,

$$1 + z \leq \frac{(1 + x_i) + (1 + x_j)}{2} = 1 + \frac{x_i + x_j}{2}.$$

Hence

$$2z \leq x_i + x_j \quad \text{and} \quad z^2 \geq x_i x_j,$$

with equality if and only if $x_i = x_j$. Applying these replacements and observations it is immediately evident that

- $M_{1,n} = c$, which is achieved if and only if one of the x_is is c and all others are 0.
- $m_{1,n} = n(\sqrt[n]{1+c} - 1)$, which is achieved if and only if $x_1 = x_2 = \cdots = x_n = \sqrt[n]{1+c} - 1$.
- $M_{2,2} = (\sqrt{1+c} - 1)^2$, which is achieved if and only if $x_1 = x_2 = \sqrt{1+c} - 1$.

For $2 \leq k \leq n$, it is evident that $m_{k,n} = 0$, which is achieved if and only if at least $n - k + 1$ of the x_is are 0.

It remains only to determine $M_{k,n}$ for $n \geq 3$ and $k \geq 2$. We may assume that at least k of the x_is are nonzero. Then $S_{k,n} > 0$, so

$$1 + c = \prod_{i=1}^{n}(1 + x_i) > 1 + \sum_{i=1}^{n} x_i, \quad \text{that is,} \quad \sum_{i=1}^{n} x_i < c \leq 1.$$

Now taking, say, $i = 1$, $j = 2$ in our earlier discussions,

$$\begin{aligned} &S_{k,n}(z, z, x_3, \ldots, x_n) - S_{k,n}(x_1, x_2, x_3, \ldots, x_n) \\ &\quad = (z^2 - x_1 x_2) S_{k-2,n-2}(x_3, \ldots, x_n) + (2z - x_1 - x_2) S_{k-1,n-2}(x_3, \ldots, x_n) \\ &\quad = (z^2 - x_1 x_2) D, \end{aligned}$$

where

$$\begin{aligned} D &= S_{k-2,n-2}(x_3, \ldots, x_n) - S_{k-1,n-2}(x_3, \ldots, x_n) \\ &\geq \left(1 - \sum_{i=1}^{n} x_i\right) S_{k-2,n-2}(x_3, \ldots, x_n) > 0. \end{aligned}$$

(We define $S_{0,n} = 1$ for $n \geq 1$.) Applying this argument repeatedly to other (x_i, x_j) pairs, and using the earlier observation that $z^2 - x_i x_j \geq 0$, we find that $M_{k,n} = \binom{n}{k}(\sqrt[n]{1+c} - 1)^k$ for $2 \leq k \leq n$ and that this is achieved if and only if $x_1 = x_2 = \cdots = x_n = \sqrt[n]{1+c} - 1$.

Also solved by Chip Curtis, Daniele Donini (Italy), and the proposer.

A Functional Inequality **October 2003**

1666. *Proposed by Razvan A. Satnoianu, City University, London, England.*

Let $f, g : [0, \infty) \longrightarrow [0, \infty)$ be functions with f increasing, g entire, and $g^{(n)}(0) \geq 0$ for all nonnegative integers n. Prove that if $x \geq y \geq z \geq 0$ (or $y \geq z \geq x \geq 0$ or $z \geq x \geq y \geq 0$), then

$$f(x)\big(g(x) - g(y)\big)(x - z) + f(y)\big(g(y) - g(z)\big)(y - x) \\ + f(z)\big(g(z) - g(x)\big)(z - y) \geq 0.$$

Solution by Daniele Donini, Bertinoro, Italy.

More generally, we prove the inequality for functions $f, g : [0, \infty) \longrightarrow [0, \infty)$ with f nondecreasing and g nondecreasing and convex. Let $x \geq y \geq z \geq 0$. If either $x = y$ or $y = z$, then the inequality follows easily, so we assume that $x > y > z \geq 0$. By the convexity of g we have

$$\frac{g(x) - g(y)}{x - y} \geq \frac{g(y) - g(z)}{y - z}.$$

Because both f and g are nonnegative and nondecreasing, it follows that

$$\big(g(x) - g(y)\big)(x - z) \geq \big(g(x) - g(y)\big)(y - z) \geq \big(g(y) - g(z)\big)(x - y) \geq 0,$$

and

$$f(x)\big(g(x) - g(y)\big)(x - z) \geq f(y)\big(g(x) - g(y)\big)(x - z) \\ \geq f(y)\big(g(y) - g(z)\big)(x - y).$$

Hence

$$f(x)\big(g(x) - g(y)\big)(x - z) - f(y)\big(g(y) - g(z)\big)(x - y) \geq 0.$$

The desired inequality now follows by adding this last result and

$$f(z)\big(g(z) - g(x)\big)(z - y) \geq 0.$$

For $y \geq z \geq x \geq 0$ or $z \geq x \geq y \geq 0$ the inequality follows by invariance under cyclic permutations of x, y, z.

Note. This problem was first published in the February 2003 issue of the MAGAZINE. However the condition on the variables was incorrectly given as $x, y, z \geq 0$ instead of $x \geq y \geq z \geq 0$. The following readers submitted counterexamples to the incorrect version: Daniele Donini (Italy), Julien Grivaux (France), Northwestern University Math Problem Solving Group, Rolf Richberg (Germany), and Li Zhou. One such example was $f(t) = t$, $g(t) = t^2$ with $x = 0$, $y = 4$, and $z = 5$.

Also solved by Michel Bataille (France), Knut Dale (Norway), Ovidiu Furdui, Elias Lampakis (Greece), Northwestern University Math Problem Solving Group, Jawad Sadek, Li Zhou, and the proposer.

Answers

Solutions to the Quickies from page 321.

A943. Because $d(f(x), f(y)) > d(x, y)$ for $x \neq y$, the function f is one to one and thus has an inverse. Let g denote the inverse of f. If $x \neq y$, then $d(x, y) =$

$d(f(g(x)), f(g(y))) > d(g(x), g(y))$, so g is continuous on X. Let $h(x) = d(x, g(x))$. Then h is a continuous real valued function on X. Because (X, d) is compact, there is an $x_0 \in X$ at which h assumes its minimum value. We claim that $g(x_0) = x_0$. Suppose not, and let $y_0 = g(x_0)$. Because $x_0 \neq y_0$ we have $d(g(x_0), g(y_0)) < d(x_0, y_0)$. But this can be rewritten as $d(y_0, g(y_0)) < d(x_0, g(x_0))$, which contradicts the minimality of $h(x_0)$. Thus, $g(x_0) = x_0$ and $x_0 = f(g(x_0)) = f(x_0)$. Because $d(f(x), f(y)) > d(x, y)$ when $x \neq y$, it is clear that f cannot have more than one fixed point.

A944. Let S_n denote the given sum. Then

$$S_n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{n-k} \binom{n+k}{n} = \sum_{k=0}^{n} a_{n-k} b_k,$$

where $a_k = (-1)^k \binom{n}{k}$ and $b_k = \binom{n+k}{n}$. Now

$$\sum_{k=0}^{n} a_k x^k = \sum_{k=0}^{n} (-1)^k \binom{n}{k} x^k = (1-x)^n$$

and

$$\sum_{k=0}^{\infty} b_k x^k = \sum_{k=0}^{\infty} \binom{n+k}{n} x^k = \frac{1}{(1-x)^{n+1}},$$

for $|x| < 1$. Thus

$$\sum_{n=0}^{\infty} S_n x^n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_{n-k} b_k \right) x^n = (1-x)^n \cdot \frac{1}{(1-x)^{n+1}} = \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n.$$

The result follows.

New Editor-Elect for the MAGAZINE

We are pleased to announce that the Board of Governors of the MAA has confirmed Allen Schwenk to serve as Editor-Elect of the MAGAZINE starting in January 2005. As of that date, new manuscripts should be sent to the Editor-Elect rather than the Editor. Details will be forthcoming in our December issue.

Schwenk is Professor of Mathematics at Western Michigan University. Michigan residents should be on the watch for MATHMAG license plates, which are rumored to be forthcoming. His term as Editor will be from 2006 to 2011.

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 mathematics literature. Readers are invited to suggest items for review to the editors.

Boutin, Chad, Purdue mathematician claims proof for Riemann hypothesis, *Purdue News* (8 June 2004), `http://news.uns.purdue.edu/UNS/html4ever/2004/040608.DeBranges.Riemann.html` . Hogan, Jenny, Mathematicians sceptical over claimed breakthrough, *New Scientist* (11 June 2004), `http://www.newscientist.com/news/news.jsp?id=ns99995104` . de Branges de Bourcia, Louis, Apology for the proof of the Riemann hypothesis, and Riemann zeta functions, both available at `http://www.math.purdue.edu/~branges/` . Conrey, J.B., and Xian-Jin Li. A note on some positivity conditions related to zeta and L-functions. *International Mathematics Research Notices* 2000 (18) 929–940 `http://arxiv.org/abs/math.NT/9812166` .

Louis de Branges de Bourcia (Purdue) has announced a proof of the Riemann hypothesis, one of the Clay Foundation's Millennium Million-Dollar Prize Problems, following a research program that he announced in 1986. However, Conrey and Li "indicate the difficulty of approaching the Riemann hypothesis by using de Branges' positivity conditions," a result that suggests that de Branges's approach can't work. (Of course, unless there's an easy proof that everybody missed, any approach will be difficult.) Nevertheless, in 1985 de Branges proved the Bieberbach conjecture, a famous unsolved problem open since 1916, about coefficients of power series of conformal maps of the unit disk.

Weisstein, Eric W., Twin prime proof proffered, *MathWorld* Headline News (9 June 2004) `http://mathworld.wolfram.com/news/2004-06-09/twinprimes/` . Arenstorf, R.F., There are infinitely many prime twins (26 May 2004) `http://arXiv.org/abs/math.NT/0405509` . Tenenbaum, G., Re: Arenstorf's paper on the twin prime conjecture, `http://listserv.nodak.edu/scripts/wa.exe?A2=ind0406&L=nmbrthry&F=&S=&P=1119` .

"In spring, a young man's fancy turns to"—announcing proofs of famous conjectures, perhaps. In May, R.F. Arenstorf (Vanderbilt University) issued a purported proof of the twin prime conjecture, via methods from analytic number theory, including a Tauberian theorem; but G. Tenenbaum (Institut Élie Cartan) quickly found a gap.

Cipra, Barry, Proof promises progress in prime progressions, *Science* 304 (2004) 1095. Peterson, Ivars, Progressive primes, `http://www.maa.org/mathland/mathtrek_04_26_04.html` (26 April 2004). Green, B., and T. Tao, The primes contain arbitrarily long arithmetic progressions, `http://arxiv.org/abs/math.NT/0404188` .

Ben Green (Pacific Institute of the Mathematical Sciences, Vancouver) and Terence Tao (UCLA) have offered a proof that there are arbitrarily long arithmetic progressions of primes. Their work, like that of de Branges and of Arenstorf, faces scrutiny before we will know if their (nonconstructive) approach has settled the question. (Cipra's title wins the headline writers' alliteration contest, though I might have gone with "Profs produce purported proof of prime progressions." In any case, I am running out of different ways to announce these proof profferings. . . .)

Collins, Graham P., The shapes of space, *Scientific American* 291 (1) (July 2004) 94–103.

In 2002 and 2003, Grigori Perelman (Steklov Institute) posted papers that—except for a minor step yet to be proved—would settle the Poincaré conjecture in the positive. His work is at a "more mature" stage of investigation by other mathematicians, who so far have "no real doubts" about it. This popular article, full of illustrations, will reintroduce readers to the lure of topology; it even mentions in passing "curious connections" to the physics of of electromagnetic interactions, to general relativity, and to string theory.

Klarreich, Erica, Theorems for sale: An online auctioneer offers math amateurs a backdoor to prestige, *Science News* 165 (12 June 2004) 376–377. Monastersky, Richard, Thumbing his nose at academe, a scholar tries to auction his services, *Chronicle of Higher Education* (28 May 2004), `http://chronicle.com/free/v50/i38/38a01501.htm` . Tozier, William, Decrease your Erdős number to 5! Scientific researcher for hire. `http://cgi.ebay.com/ws/eBayISAPI.dll?ViewItem&item=3189039958` ; Notional slurry [Tozier's blog about the auction], `http://williamtozier.com/slurry/comment/social/erdos6.html` . Grossman, Jerry, et al., The Erdős number project. `http://www.oakland.edu/enp/` .

Your *Erdős number* is your degree of separation from the late Paul Erdős in the graph connecting mathematicians who have written joint research papers (technically, the distance in edges of the shortest path from you to Erdős in the mathematics research collaboration graph). William Tozier, a scientist with Erdős number 4, "just for fun" recently auctioned on the Internet market eBay 40 hours of his time for collaboration on a joint research paper. At least 100 people showed interest, but the winning bidder (at $1,031) was a saboteur: José Burillo (Polytechnic University of Barcelona), who already has an Erdős number of 3, refuses to pay up or to collaborate with Tozier and bid just to stop the "travesty." Tozier has been inspired by the responses to try to create a "collaborative community of amateur mathematicians." (More than 160,000 mathematicians have an Erdős number of 5 or less. Maybe you do, but I don't, so all this emphasis on Erdős numbers makes me feel a bit left out, especially since the Combined Membership List of the American Mathematical Society and other societies (including the Mathematical Association of America) has only 52,000 names. But if it were important for me to have an even lower Erdős number than Tozier's, I could approach my colleague in the next office, who can be neither blandished nor bribed; but I doubt that I could rise to a high enough fraction of his level of talent to merit co-authorship.) Meanwhile, Erdős's death in 1996 has only slowed his output of 1,500 publications; he continues through mathematicians finally writing up joint work that they did with him a decade or so ago.

Bartlett, Albert A., et al., *The Essential Exponential! For the Future of Our Planet*, Center for Science, Mathematics & Computer Education, University of Nebraska—Lincoln, 2004; 291 pp, $25 (P); 4 or more copies to the same address, $12.50 each, postpaid. ISBN 0–9758973–0–6. Case, James, Meeting the world's energy needs, *SIAM News* 37 (5) (June 2004) 4, 9.

"The greatest shortcoming of the human race is our inability to understand the exponential function." This is the opening line of a talk that Al Bartlett (retired professor of physics, University of Colorado, and former president of the American Association of Physics Teachers) has been giving for 35 years (1,500 times!). This book reprints articles of his on the energy crisis, population, resources, the arithmetic of growth, and the exponential function, together with a few related articles by others, including M. King Hubbert. Hubbert is famous for his prediction in the 1950s that U.S. oil production follows a Gaussian distribution and would peak ("Hubbert's peak") in 1970 (which it did). Author Case reviews two other books on modeling the depletion of fossil fuels.

Gielis, Johan, A generic geometric transformation that unifies a wide range of natural and abstract shapes, *American Journal of Botany* 90 (3) (2003) 333–338

Gielis generalizes superellipses ($(x/a)^n + (y/b)^n = 1$) to shapes with arbitrary rotational symmetry, using trigonometric functions and Fourier series. He cites other articles in which he models biological forms via his "Superformula."

NEWS AND LETTERS

Carl B. Allendoerfer Award — 2004

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in *Mathematics Magazine*. The Awards are named for Carl B. Allendoerfer, a distinguished mathematician at the University of Washington and President of the Mathematical Association of America, 1959–60.

Charles I. Delman and Gregory Galperin, A Tale of Three Circles, MATHEMATICS MAGAZINE, February 2003, pp.15–32.

The article by Charles Delman and Gregory Galperin begins with an intriguing basic question about the sum of the angles of curvilinear triangles formed by the arcs of three circles in the plane. In the course of analyzing the problem, the authors carry us along a wave that takes us through examples, a theorem that explains it all, and an overview of three classical geometries. The authors consider three configurations of three intersecting circles in the plane: first, the case where the three circles intersect at a common point and no circles are tangent to each other; next, the case where the three circles have collinear centers; and finally, the case in which the three circles intersect as in a generic Venn diagram. Each of the three cases results in a different sum of the angles of a curvilinear triangle. A very interesting paper so far, but the fun is just beginning. The authors open a window with the basic question and lead us to a panoramic view of noneuclidean geometries. With careful summaries of spherical and hyperbolic geometries and an introduction to stereographic projection, the authors succeed masterfully in sharing the beauty and fascination of noneuclidean geometries with those unfamiliar with these geometries. The new perspectives and the proof linking the three geometries to the three configurations of three intersecting circles is an example of elegant mathematics written in an accessible and clear style.

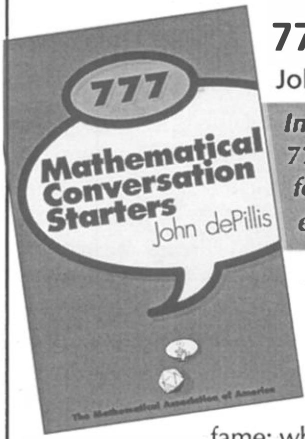
Biographical Note: Charles I. Delman Charles Delman, currently Professor of Mathematics at Eastern Illinois University, grew up in Manhattan. New York City's rich multi-cultural environment led to lifelong loves for—and dabbling in—art, music, and ethnic food, while trips to the nearby Museum of Natural History and summers in the Catskill mountains cultivated his passions for nature and science. He enjoys backpacking with his children, Anna and Ben, and his partner, Barbara. He is also committed to political activism for environmental preservation, peace, and social justice. Charles received his bachelor's in mathematics from Harvard and his Ph.D. from Cornell, under the guidance of Alan Hatcher, to whom he will always be grateful for demonstrating so well how to get at the essence of an idea. Before coming to EIU, he taught at The Ohio State University and Pitzer College. His mathematical interests include low-dimensional topology, classical geometry, and dynamical systems.

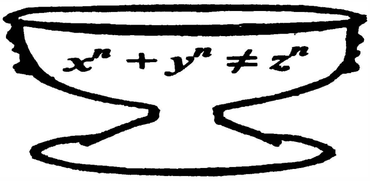
Response from Charles Delman I am deeply honored and excited to receive the Allendoerfer Prize. Writing "A Tale of Three Circles" was a great pleasure, and I am greatly indebted to the people who enhanced both that pleasure and the quality of the article: the editor, Frank Farris, who kept us striving for greater clarity and liveliness with his many constructive criticisms, questions, and suggestions; my delightful geometry students, who have over the years so greatly stimulated my interest in the subject; my loving partner, Barbara Lawrence, who puts up with more than can be mentioned; and, of course, my co-author, Gregory Galperin, with whom I have shared many pleasurable hours of inquiry and collaboration, and who asked in the first place the innocent little question that led to this whole thing.

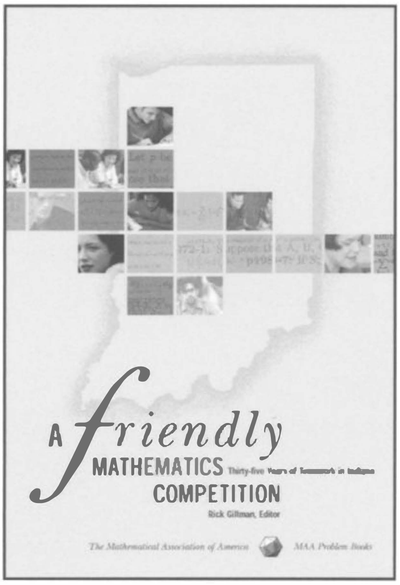
Biographical Note: Gregory Galperin Gregory Galperin, currently Professor of Mathematics at Eastern Illinois University, was born in Tbilisi, the capital of Georgia, USSR (Georgia is now a separate country). At age 14, he became a student of the famous A.N.Kolmogorov physics/mathematics school in Moscow (USSR), and later on a student of the physical and mathematical Department at the University of Moscow. He received his PhD from the University of Moscow under the tutelage of prominent twentieth century mathematician Andrei N. Kolmogorov, to whom he will always be grateful for demonstrating to Dr. Galperin the diversity of mathematics and illustrating the elegantly simple yet profound ideas that connect various branches of mathematics. Dr. Galperin's Ph.D. thesis concentrated on dynamical systems with local interaction, which arose partially from biology and partially from automata theory. Later on, he worked with Prof. Ya. G. Sinai (Moscow University and Princeton University) on the theory of billiards. He has published more than 50 mathematical articles on billiards and other dynamical systems, on combinatorial geometry, on differential geometry, and on celestial mechanics. Dr. Galperin has been an Alexander von Humboldt fellow since 1994, and has collaborated with Prof. S. Albeverio (the University of Bonn, Germany). As an undergraduate, Dr. Galperin was involved in mathematical olympiads, mainly as a creator of new mathematical problems. He helped conduct the Moscow Mathematical Olympiads and the Russian National Math Olympiads from 1970–1980, and in 1986 he published the book *Moscow Mathematical Olympiads*. He wrote and published many popular articles and problems on mathematics in the journals *Kvant* (in Russian) and *Quantum* (in English). He has also played a big role in conducting the American National Olympiads, the USAMO, since 1996, has served as Coordinator at the 42nd International Mathematical Olympiad in Washington, D.C., and was the Deputy Leader of the USA team at the 44th International Mathematical Olympiad in Japan, 2003.

Gregory Galperin plays table tennis, likes to draw and listen to music, and enjoys reading literature.

Response from Gregory Galperin Being awarded the Allendoerfer Prize comes as an unexpected honor and a very delightful surprise for me. In all of my mathematical investigations, I have always tried to express the beauty and uncommonness of that one key idea which resolves an initially impenetrable challenge. Such an idea could consist of just a single unusual expression or word, or an unorthodox mathematical construction, or even an unusual mathematical theory. In my common article with C. Delman the initial challenge consisted of finding the angle sum of the curvilinear triangle formed by three upper semicircles with collinear centers, and noticing the hyperbolic geometry behind the picture turned out to be the key idea that solved the problem. Inspired by this unusual association, and having long aspired to reveal the unity of three famous geometries—the Euclidean, spherical, and hyperbolic—to a broad audience, we materialized our dream through a narrative of the workings of the three geometries in a standard sphere. In the course of our many years of work on the article, we both enjoyed seeing our project materialize, as well as having the opportunity to share the beauty of the geometrical ideas we used with numerous other lovers of mathematics. We presented our article at various workshops and colloquia, to students and professors of various universities, and only after much long-term polishing did we submit our final text to MATHEMATICS MAGAZINE. I kindly thank everyone who gave me good advice while work was being done on the article, and I especially thank the MAA for its deep appreciation of my collaborative efforts with Dr. Delman, my coauthor and friend.


777
Mathematical Conversation Starters
John dePillis
The Mathematical Association of America


$x^n + y^n \neq z^n$


A friendly
MATHEMATICS
Thirty-five Years of Teamwork in Indiana
COMPETITION
Rick Gillman, Editor
The Mathematical Association of America
MAA Problem Books

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, DC 20036