Vol. 81, No. 4, October 2008

MATHEMATICS MAGAZINE

$141 = 3 \cdot 47$	$39693 = 3 \cdot 101 \cdot 131$	$14273478 = 2 \cdot 3^2 \cdot 23^2 \cdot 1499$
$142 = 2 \cdot 71$	$39694 = 2 \cdot 89 \cdot 223$	$14273479 = 11 \cdot 67 \cdot 107 \cdot 181$
$143 = 11 \cdot 13$	$39695 = 5 \cdot 17 \cdot 467$	$14273480 = 2^3 \cdot 5 \cdot 13 \cdot 27449$
$144 = 2^4 \cdot 3^2$	$39696 = 2^4 \cdot 3 \cdot 827$	$14273481 = 3 \cdot 29 \cdot 359 \cdot 457$
$145 = 5 \cdot 29$	$39697 = 7 \cdot 53 \cdot 107$	$14273482 = 2 \cdot 113 \cdot 137 \cdot 461$
$146 = 2 \cdot 73$	$39698 = 2 \cdot 23 \cdot 863$	$14273483 = 7 \cdot 53 \cdot 79 \cdot 487$
$147 = 3 \cdot 7^2$	$39699 = 3^2 \cdot 11 \cdot 401$	$14273484 = 2^2 \cdot 3 \cdot 19 \cdot 62603$
$148 = 2^2 \cdot 37$	$39700 = 2^2 \cdot 5^2 \cdot 397$	$14273485 = 5 \cdot 31 \cdot 71 \cdot 1297$

Consecutive Integers with Equally Many Principle Divisors

- The Shot Made Round the Table
- Probabilistic Reasoning Is Not Logical
- Wazir Circuits on an Obstructed Chessboard

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 Farris, Editor, Mathematics Magazine, Department of Mathematics, Santa Clara University, 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 credit: The cover image illustrates consecutive integers with two principle divisors, with three principle divisors, and with four principle divisors. How long a run can we have with equally many principle divisors? Our lead article examines this and related questions.

AUTHORS

Roger Eggleton is Professor of Mathematics at Illinois State University, where he convenes the Discrete Mathematics group, is advisor for the Math Club, and coaches the Putnam team. He received his PhD at University of Calgary (1973), with Richard Guy as advisor. He has published over sixty research papers, mainly in number theory, graph theory, and combinatorics. Among them are fourteen joint papers with Jim MacDougall. His wide-ranging other interests include active involvements with languages and philately.

Jim MacDougall has been enjoying teaching mathematics and doing research in graph theory and number theory at the University of Newcastle in Australia for 20 years, after doing the same for 18 years at University of Prince Edward Island in Canada. In 2004 Jim completed a PhD under the supervision of longtime colleague and co-author Roger Eggleton. Jim also finds time to enjoy choral singing and stamp collecting.

Ruma Falk is professor emerita of psychology at the Hebrew University of Jerusalem, where she studied mathematics, statistics, and psychology and completed her PhD thesis on the perception of randomness (supervised by Amos Tversky) in 1975. She combined her love of mathematics and psychology (no contradiction between the two) in her teaching and many research projects on statistical thinking, judgment of randomness and coincidences, and children's concepts of probability and infinity. She finds herself time and again entangled in solving problems, puzzles, and paradoxes. Otherwise, she is addicted to the music of Johann Sebastian Bach.

John Riley has been a professor at Bloomsburg University of Pennsylvania since 1984. He teaches mathematics and computer forensics. Currently he is the chair of the university wide curriculum committee and co-chair of the general education task force. His PhD is from the University of Connecticut. He also serves as President of the Bloomsburg Area School District Board of Directors. In his spare time he can be found running in and out of the Susquehanna River Valley.

Irl Bivens has taught at Davidson College since 1982. He did his undergraduate work at Pfeiffer College and his graduate work at the University of North Carolina at Chapel Hill. His thesis advisor was Patrick Eberlein. Irl's mathematical interests include differential geometry, recreational mathematics, and the history of mathematics. Since the inception of the North Carolina State Mathematics Contest thirty years ago, he and his colleague Ben Klein have written nearly half of the contests.

Ben Klein is Dolan Professor of Mathematics at Davidson College. He began teaching at Davidson in 1971 and will retire at the end of the 2007–08 academic year. He did his undergraduate work at the University of Rochester and his graduate work at Yale University. His thesis advisor was G. A. Hedlund. Ben's graduate work and his early interests were in analysis, but, over the years, his interests have expanded to include probability and combinatorics. He is active in the Advanced Placement Calculus Program and attended his 18th consecutive Calculus Reading in June 2008.

Arthur Holshouser is an amateur, in the true and very best sense of the word, mathematician who lives in Charlotte, North Carolina. His mathematical interests are wide-ranging, and he has worked and published in a number of different fields, including combinatorial games. His most recent interests focus on the interplay between algebra and combinatorics. He collaborates with faculty at both Davidson College and the University of North Carolina at Charlotte and is a regular contributor to the North Carolina State Mathematics Contest.

Vol. 81, No. 4, October 2008

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

MAA Advertising
1529 Eighteenth St. NW
Washington DC 20036

Phone: (866) 821-1221
Fax: (202) 387-1208
E-mail: advertising@maa.org

Further advertising information can be found online at www.maa.org

Change of address, missing issue inquiries, and other subscription correspondence:

MAA Service Center, *maahq@maa.org*

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 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

ARTICLES

Consecutive Integers with Equally Many Principal Divisors

ROGER B. EGGLETON
Illinois State University
Normal, IL 61790, U.S.A.

JAMES A. MACDOUGALL
The University of Newcastle
Newcastle, N.S.W. 2308, Australia

1. Introduction

Classifying the positive integers as primes, composites, and the unit, is so familiar that it seems inevitable. However, other classifications can bring interesting relationships to our attention. In that spirit, let us classify positive integers by the number of *principal divisors* they possess, where we define a principal divisor of a positive integer n to be any prime-power divisor $p^a \mid n$ which is maximal (so p is prime, a is a positive integer, and p^{a+1} is not a divisor of n). The standard notation $p^a \,||\, n$ can be read as "p^a is a principal divisor of n."

The Fundamental Theorem of Arithmetic is usually stated in a form emphasizing how primes enter the structure of the positive integers, such as: *Every positive integer is the product of a unique finite multiset of primes.* (Recall that a multiset is a collection of elements in which multiple occurrences are permitted.) Alternatively, the Fundamental Theorem of Arithmetic can be stated in a form that focuses on how maximal prime-powers enter the structure of the positive integers, such as: *Every positive integer is the product of a unique finite set of powers of distinct primes.* Consequently every positive integer is the product of its principal divisors, and every finite set of powers of distinct primes is the set of principal divisors of a unique positive integer. Of course, the number of principal divisors of n is equal to the number of distinct prime factors of n, but here the principal divisors are the simple structural components of n, whereas the distinct prime factors are but a shadow of that structure. Readers who find the present paper of interest might find similar interest in [**6**], where upper bounds on the sum of principal divisors of n are established by elementary means.

For each integer $n \geq 0$, let P_n be the set of all positive integers with exactly n principal divisors, so $P_0 = \{1\}$, and

$$P_1 = \{2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 23, 25, 27, 29, 31, 32, 37, \ldots\},$$

$$P_2 = \{6, 10, 12, 14, 15, 18, 20, 21, 22, 24, 26, 28, 33, 34, 35, 36, 38, \ldots\},$$

$$P_3 = \{30, 42, 60, 66, 70, 78, 84, 90, 102, 105, 110, \ldots\},$$

$$P_4 = \{210, 330, 390, 420, 462, 510, \ldots\}, \text{ etc.}$$

In particular, P_1 comprises the prime-powers, or *principal* integers; P_2 comprises the products of two coprime principal integers, or *rank 2* integers; and so on. Collectively, we call $\{P_n : n \geq 0\}$ the *rank sets* of positive integers.

Clearly the rank sets are a partition of the positive integers, by the Fundamental Theorem of Arithmetic. Thus it is interesting to look at the occurrence of runs of consecutive integers within each rank set: this is one of the relationships immediately brought into focus by the classification.

For compactness, let us write $a^{[r]}$ to denote the run of r consecutive integers beginning with a, where a and r are positive integers, so

$$a^{[r]} = \{a + i : 0 \leq i < r\}.$$

We call r the *size* of the run. In particular, the run $a^{[r]}$ is *nontrivial* if $r \geq 2$, and $a^{[r]} \subset P_n$ is a *maximal* run in P_n if it is nontrivial and P_n contains neither $a - 1$ nor $a + r$. Thus, the first few maximal runs in P_1 are $2^{[4]}$, $7^{[3]}$, $16^{[2]}$, $31^{[2]}$, $127^{[2]}$ and $256^{[2]}$; the first few maximal runs in P_2 are $14^{[2]}$, $20^{[3]}$, $33^{[4]}$, $38^{[3]}$, $44^{[3]}$, $50^{[3]}$ and $54^{[5]}$.

It is easy to see that the size of runs in P_n is bounded. For if M is the product of the first $n + 1$ primes, then $M \in P_{n+1}$ and any run of M consecutive integers contains a multiple of M, so any run in P_n has size less than M. Thus, for each integer $n \geq 0$ there is a positive integer $r(n)$ which is the *maximum size* attained by runs in P_n. Trivially, $r(0) = 1$. We have already seen that $r(1) \geq 4$ and $r(2) \geq 5$, and we shall soon see that in fact $r(1) = 4$. Our main objective is to study $r(2)$, which we shall determine "within 1". Later we shall also discuss $r(n)$ for $n \geq 3$.

2. Maximal runs in P_1 and P_2

Returning to the principal integers, it is clear that any nontrivial run in P_1 contains an even integer so, being principal, any such integer must be a power of 2. Since 2 and 4 are the only powers of 2 that differ by 2, any maximal run of principal integers greater than 5 must contain exactly one even integer, so $2^{[4]}$ is the unique longest run in P_1, and $r(1) = 4$.

The long-standing conjecture credited to Catalan, that 8 and 9 are the only two consecutive integers which are nontrivial prime powers, was recently proved by Mihailescu [**11**]. From this it follows that any maximal run of principal integers greater than 9 must contain a power of 2 and any adjacent number in the run must be a prime. It is well known that $2^n - 1$ can only be prime when n itself is prime, and $2^n + 1$ can only be prime when n is a power of 2: when $n \geq 3$ these two conditions are mutually exclusive, so any maximal run of principal integers greater than 9 has just two members.

Primes of the form $2^n + 1$ are *Fermat primes*. The only known Fermat primes are 3, 5, 17, 257 and 65537, but no proof is known that there are no others. Primes of the form $2^n - 1$ are *Mersenne primes*. Currently 44 Mersenne primes are known [**7**]. A distributed computing project known as GIMPS (the Great International Mersenne Prime Search) has made numerous additions to this list in recent years, but no proof is known that infinitely many such primes exist. Consequently, although we know $2^{[4]}$ and $7^{[3]}$ are the only maximal runs of more than two principal integers, it is not known whether P_1 contains infinitely many runs of size 2.

What happens with maximal runs of rank 2 numbers? This is less familiar territory, so one does not know quite what to expect. We shall prove:

THEOREM 1. *There is no run of* 10 *consecutive integers in* P_2.

Hence $r(2) \leq 9$. Is this a "sharp" result? It turns out to be "within 1" of the exact value of $r(2)$. Our methods appear unable to decide on the existence of runs of size 9 in P_2, but our results strongly suggest the following:

CONJECTURE 1. *In P_2 there is no run of size 9 and the only maximal runs of size greater than 6 are $141^{[8]}$, $212^{[8]}$, $323^{[7]}$ and $2302^{[7]}$.*

It may also be true that $91^{[6]}$ is the only maximal run of size 6 in P_2, but we have less information about runs of size 6 than about longer runs.

3. Størmer's theorems

For any set of primes P let $S(P)$ be the set of positive integers with all their prime factors in P, that is, $S(P)$ is the multiplicative semigroup of positive integers generated by P. Some results about the integers in $S(P)$ turn out to be among the main tools we require to prove Theorem 1.

If D is a non-square integer, the Pell equation $x^2 - Dy^2 = c$ always has positive integer solutions when $c = 1$ (see **[12]**, for example). However, for other values of c it need not have positive integer solutions; for instance, this is the case when $c = -1$ if $D \equiv 3 \pmod 4$. But if $c \in \{1, -1\}$ and $x^2 - Dy^2 = c$ does have positive integer solutions, then all such solutions are generated by the smallest positive solution (x_0, y_0), usually called the *fundamental solution* or *minimal solution* **[12, 16]**. In 1897 Carl Størmer published the following theorem **[13]** which identifies a remarkable property of the fundamental solution.

STØRMER'S PELL EQUATION THEOREM. *Let D be a non-square positive integer, let P be the set of prime divisors of D, and let $c \in \{1, -1\}$. If the Pell equation $x^2 - Dy^2 = c$ has positive integer solutions and (x_0, y_0) is the fundamental solution, then y_0 is the only solution for y that is a possible member of $S(P)$.*

For example, if $D = 2$ then $S(P)$ comprises all the powers of 2. Thus the fundamental solution $(x, y) = (3, 2)$ is the only solution to $x^2 - 2y^2 = 1$ in positive integers with y a power of 2; similarly $(1, 1)$ is the only solution to $x^2 - 2y^2 = -1$ in positive integers with y a power of 2.

Størmer used his theorem to prove the following result **[14]**, first published in 1898.

STØRMER'S NEIGHBORING PAIRS THEOREM. *For given positive integers A, B, m, n, $a_1, \ldots, a_m$, $b_1, \ldots, b_n$, there are at most finitely many sequences of positive integers $x_1, \ldots, x_m$, $y_1, \ldots, y_n$ such that*

$$|Aa_1^{x_1} \cdots a_m^{x_m} - Bb_1^{y_1} \cdots b_n^{y_n}| \leq 2.$$

All solutions follow from the fundamental solutions to a finite set of Pell equations determined by A, B, m, n, $a_1, \ldots, a_m$, $b_1, \ldots, b_n$.

In principle it is routine to set up the Pell equations referred to in the Neighboring Pairs Theorem, and to determine the fundamental solutions of those that do have solutions, so Størmer effectively provided an algorithm for solving the Diophantine inequality in the Neighboring Pairs Theorem. Subsequently he gave simpler proofs of both theorems in a paper **[15]** that we commend to the reader. We shall soon solve two special cases of the neighboring pairs problem by methods illustrating Størmer's arguments: the results are stated below as a Corollary to the Neighboring Pairs Theorem.

If P is finite, then by Størmer's Neighboring Pairs Theorem there are only finitely many pairs of consecutive integers in $S(P)$. D. H. Lehmer **[9]** gave a new proof of

Størmer's theorem for the case of consecutive integers, and explicitly computed the last pairs of consecutive integers in $S(P)$ when P is any initial subset of the primes with largest member 41 or less [**10**]. Some related tabulations are given in [**4**], and some elementary arguments establishing special instances of Størmer's result are given in [**5**] and [**8**]. The latter paper, by Halsey and Hewitt, discusses the fascinating connection between fundamental frequency ratios in Western music and consecutive pairs of integers in $S(2, 3, 5)$.

TABLE 1: Last maximal runs in $S(P)$

$S(P)$	Last maximal runs
$S(2,3)$	$8^{[2]}, 2^{[3]}$
$S(2,3,5)$	$80^{[2]}, 8^{[3]}, 3^{[4]}$
$S(2,3,5,7)$	$4374^{[2]}, 48^{[3]}, 7^{[4]}$

For our present purposes, we note in TABLE 1 the last maximal runs of various sizes in $S(P)$ when P is an initial subset of the primes with largest member 7 or less. We shall also need the following consequence of Størmer's theorem:

COROLLARY 1. *The only pairs of integers satisfying* $|3^a - 5^b| = 2$ *are* $\{3, 5\}$ *and* $\{25, 27\}$, *and the only pair satisfying* $|3^a - 7^b| = 2$ *is* $\{7, 9\}$.

Proof. Størmer's method applied to pairs of integers $\{3^a, 5^b\}$ satisfying $|3^a - 5^b| = 2$ puts them equal to $x - 1$ and $x + 1$, so their product is of the form $x^2 - 1 = Dy^2$, where $y \in S(3,5)$ and $D \in \{3^a 5^b : 1 \le a, b \le 2\}$ is non-square, so $D \in \{15, 45, 75\}$. For these three values of D, the corresponding Pell equations have fundamental solutions $(x_0, y_0) = (4, 1), (161, 24)$ and $(26, 3)$ respectively. Now Størmer's Pell Equation Theorem shows that $y_0 = 1$ and 3 are the only y solutions of these Pell equations that lie in $S(3,5)$. The corresponding x solutions yield the pairs $\{x_0 - 1, x_0 + 1\} = \{3, 5\}$ and $\{25, 27\}$, so these are the only pairs $\{3^a, 5^b\}$ that differ by 2.

Similarly, pairs of integers $\{3^a, 7^b\}$ satisfying $|3^a - 7^b| = 2$ must correspond to fundamental solutions of $x^2 - 1 = Dy^2$, where $y \in S(3,7)$ and D is non-square, so $D \in \{21, 63, 147\}$. The corresponding Pell equations have fundamental solutions $(x_0, y_0) = (55, 12), (8, 1)$ and $(97, 8)$ respectively. Now $y_0 = 1$ is the only y solution of these Pell equations that lies in $S(3,7)$. The corresponding x solution yields the pair $\{x - 1, x + 1\} = \{7, 9\}$, so this is the only pair $\{3^a, 7^b\}$ differing by 2. ∎

4. Constraints on runs in P_2

This section builds a proof of Theorem 1. Our method is to establish several properties of runs in P_2 that allow us to close in on the possible structure of long runs. Finally we accumulate enough constraints to show that no run of size 10 or more could possibly satisfy all the constraints, so we can conclude that every run in P_2 has size less than 10. Consequences of Størmer's Neighboring Pairs Theorem are used in several key steps, including some of the results in TABLE 1 and our Corollary to the Neighboring Pairs Theorem.

PROPERTY 1. *Any run of consecutive integers in* P_2 *contains at most one multiple of* 6.

Proof. Since 89 and 97 are the first two consecutive primes with difference greater than 6, it follows that between any two consecutive multiples of 6 less than 96 there is at least one prime. Since primes are in P_1, no run in P_2 can contain two consecutive multiples of 6 less than 96. On the other hand, 8 and 9 are the last two consecutive integers in $S(2, 3)$, so 48 and 54 are the last two consecutive multiples of 6 in P_2. The property follows. ∎

The next three properties concern multiples of 5 in P_2 that occur within runs which contain a multiple of 6.

PROPERTY 2. *Only one maximal run of consecutive integers in P_2 contains a multiple of 6 and a multiple of 5 which differ by 2 or 3; that run is $158^{[5]}$.*

Proof. If $6a$ and $5b$ are members of P_2 that differ by 2, then $6a \in S(2, 3)$ and $5b = 10c \in S(2, 5)$ for some integer c. Then $3a$ and $5c$ are consecutive integers in $S(2, 3, 5)$, and $80^{[2]}$ is the last such pair. Hence the corresponding pairs $\{6a, 5b\}$ in P_2 are $\{10, 12\}$, $\{18, 20\}$, $\{48, 50\}$ and $\{160, 162\}$. For the first three pairs, the intervening number is in P_1; hence the only maximal run in P_2 that contain integers $6a$ and $5b$ with $|6a - 5b| = 2$ is $158^{[5]}$.

Similarly, if $6a$ and $5b$ are members of P_2 that differ by 3, then $6a \in S(2, 3)$ and $5b = 15c \in S(3, 5)$ for some integer c, so $2a$ and $5c$ are consecutive integers in $S(2, 3, 5)$. The corresponding pairs $\{6a, 5b\}$ in P_2 are $\{12, 15\}$, $\{15, 18\}$, $\{45, 48\}$ and $\{72, 75\}$, but in each case there is an intervening prime, so no maximal run in P_2 contains integers $6a$ and $5b$ with $|6a - 5b| = 3$. ∎

PROPERTY 3. *No run of consecutive integers in P_2 contains a multiple of 6 and a multiple of 5 which differ by 4.*

Proof. First suppose $6a$ and $20b$ are members of P_2 that differ by 4, so $6a = 12c \in S(2, 3)$ for some integer c, and $20b \in S(2, 5)$. Then $5b$ and $3c$ are consecutive integers in $S(2, 3, 5)$. Since $80^{[2]}$ is the last such pair, the corresponding pairs $\{6a, 20b\}$ in P_2 are $\{20, 24\}$, $\{36, 40\}$, $\{96, 100\}$ and $\{320, 324\}$. The first three pairs have an intervening prime, while the last pair is not in a run in P_2 because $322 \in P_3$. Thus, no run of consecutive integers in P_2 contains a multiple of 6 and a multiple of 20 differing by 4.

Now suppose $6a$ and $10d$ are members of P_2 that differ by 4, and d is odd. Then $|3a - 5d| = 2$, so a is also odd. But $6a \in S(2, 3)$ and $10d \in S(2, 5)$, so a is a power of 3 and d is a power of 5. By Corollary 1, $\{3, 5\}$ and $\{25, 27\}$ are the only pairs of proper powers of 3 and 5 that differ by 2. The corresponding pairs $\{6a, 10d\}$ in P_2 are $\{6, 10\}$ and $\{50, 54\}$, but each has an intervening prime, so neither pair is contained in a run in P_2. The property follows. ∎

It is noteworthy that the pair $\{320, 324\}$, appearing in the proof of Property 3, actually corresponds to a near miss: P_2 contains the two maximal runs $319^{[3]}$ and $323^{[7]}$, and the only intervening integer is $322 \in P_3$. The two bordering integers are $318 \in P_3$ and $330 \in P_4$ (both multiples of 6), and their neighbors 317 and 331 are consecutive primes.

PROPERTY 4. *Any run of consecutive integers in P_2 contains at most one multiple of 5.*

Proof. On the contrary, suppose there is a run of consecutive integers in P_2 that contains two multiples of 5. Let R be that portion of the run which begins and ends with two consecutive multiples of 5. Since R has size 6, we have $6a \in R$ for some integer a. But $6a \in P_2$ so it is distinct from the multiples of 5. If $6a$ differs by 2 from the nearer multiple of 5, these two members of R belong to the maximal run $158^{[5]}$, by Property 2. But this does not contain two multiples of 5, so is disjoint from R. Hence

$6a$ must be adjacent to the nearer multiple of 5. But then it must differ by 4 from the other multiple of 5 in R, and Property 3 shows that no run in P_2 contains two such numbers. Hence, by contradiction, R does not exist. ■

Since any run of 10 consecutive integers contains two multiples of 5, Property 4 immediately implies our target result:

THEOREM 1. *There is no run of* 10 *consecutive integers in* P_2.

COROLLARY 2. *Runs in* P_2 *have maximum size* $r(2) = 8$ *or* 9.

Although our methods do not seem to be strong enough to decide whether runs of size 9 exist in P_2, consideration of multiples of 7 yields further properties, revealing more of the structure of P_2. In particular, we are led to discover the examples of runs of size 8 in P_2 which are incorporated in Conjecture 1. We shall pursue this in the next section.

5. Further constraints on runs in P_2

Now that we know P_2 has no runs of size 10 or more, the study of long runs of rank 2 integers can proceed by asking: What is the structure of any run of size 7 or more in P_2? Since any run of 7 consecutive integers must contain a multiple of 7, at least one multiple of 6, and at least one multiple of 5, the relative placement of these multiples will now be considered.

PROPERTY 5. *Only two maximal runs of consecutive integers in* P_2 *contain a multiple of* 6 *and a multiple of* 7 *which differ by* 2 *or* 3 : *those runs are* $54^{[5]}$ *and* $141^{[8]}$.

Proof. If $6a$ and $7b$ are members of P_2 that differ by 2, then $6a \in S(2, 3)$ and $7b = 14c \in S(2, 7)$ for some integer c. Then $3a$ and $7c$ are consecutive integers in $S(2, 3, 7)$. From the fact that $4374^{[2]}$ is the last pair in $S(2, 3, 5, 7)$, a straightforward calculation verifies that $63^{[2]}$ is the last pair in $S(2, 3, 7)$. The corresponding pairs $\{6a, 7b\}$ in P_2 are $\{12, 14\}$, $\{54, 56\}$ and $\{96, 98\}$. For the first and third pair, the intervening number is prime; hence the only maximal run in P_2 that contains integers $6a$ and $7b$ with $|6a - 7b| = 2$ is $54^{[5]}$.

Similarly, if $6a$ and $7b$ are members of P_2 that differ by 3, then $6a \in S(2, 3)$ and $7b = 21c \in S(3, 7)$ for some integer c, so $2a$ and $7c$ are consecutive integers in $S(2, 3, 7)$. The corresponding pairs $\{6a, 7b\}$ in P_2 are $\{18, 21\}$, $\{21, 24\}$, $\{81, 84\}$, $\{144, 147\}$ and $\{189, 192\}$. For all but one of these pairs, there is an intervening prime; hence the only maximal run in P_2 that contains integers $6a$ and $7b$ with $|6a - 7b| = 3$ is $141^{[8]}$. ■

PROPERTY 6. *Any run of consecutive integers in* P_2 *contains at most one multiple of* 7.

Proof. On the contrary, suppose there is a run R of consecutive integers in P_2 that contains two consecutive multiples of 7. Since they are in P_2, neither is a multiple of 6, so there is a multiple of 6 between them but not adjacent to either of them. Thus the multiple of 6 and the nearer multiple of 7 differ by 2 or 3, so R must be contained in $54^{[5]}$ or $141^{[8]}$, by Property 5. But each of these runs contains only one multiple of 7, so by contradiction it follows that R does not exist. ■

PROPERTY 7. *There are no runs of consecutive integers in* P_2 *that contain a multiple of* 6 *and a multiple of* 7 *which differ by* 4.

Proof. First suppose $6a$ and $28b$ differ by 4 and belong to some run of consecutive integers in P_2. Then $6a = 12c \in S(2,3)$ for some integer c, and $28b \in S(2,7)$, so $7b$ and $3c$ are consecutive integers in $S(2,3,7)$. As noted in the proof of Property 5, $63^{[2]}$ is the last pair in $S(2,3,7)$. The corresponding pairs in P_2 are $\{24, 28\}$, $\{108, 112\}$ and $\{192, 196\}$, but none of them is contained in a run in P_2.

Now suppose, for some odd integer d, that $6a \in S(2,3)$ and $14d \in S(2,7)$ differ by 4 and belong to some run of consecutive integers in P_2. Then $|3a - 7d| = 2$, so a is odd; hence $3a$ is a power of 3 and $7d$ is a power of 7. By our corollary to the Neighboring Pairs Theorem, $\{7, 9\}$ is the only pair of proper powers of 3 and 7 that differ by 2; the corresponding pair $\{6a, 14d\} = \{14, 18\}$ in P_2 is not contained in a run in P_2. The property follows. ∎

PROPERTY 8. *Exactly two maximal runs of consecutive integers in P_2 contain a multiple of* 6 *and a multiple of* 35: *these are* $33^{[4]}$ *and* $4374^{[2]}$.

Proof. If $6a$ and $35b$ are members of some run of consecutive integers in P_2, then $6a \in S(2,3)$ and $35b \in S(5,7)$, so $|6a - 35b|$ cannot be a multiple of 2, 3, 5 or 7. But every run in P_2 has size less than 10, by our Theorem, so $|6a - 35b| = 1$, and $6a$ and $35b$ are consecutive integers in $S(2,3,5,7)$. Since $4374^{[2]}$ is the last nontrivial run in $S(2,3,5,7)$, the corresponding pairs in P_2 are $\{35, 36\}$ and $\{4374, 4375\}$. The corresponding maximal runs in P_2 are $33^{[4]}$ and $4374^{[2]}$. ∎

PROPERTY 9. *Any run of size at least* 7 *in P_2 contains exactly one multiple of* 6, *exactly one multiple of* 5, *and exactly one multiple of* 7, *and these are three distinct members of the run. The multiple of* 5 *is always adjacent to the multiple of* 6. *If the run has size* 8 *or more, the multiple of* 7 *is also adjacent to the multiple of* 6, *except in the case of the maximal run* $141^{[8]}$.

Proof. Uniqueness of the multiples of 6, 5 and 7 follows from Properties 1, 4 and 6 respectively. By Property 8, the only two runs in P_2 containing a multiple of 6, and a multiple of 5 which is also a multiple of 7, have size less than 7. Hence, in any run of size 7 or more, all three must be distinct integers. By Properties 2 and 3, in any run of size at least 6 in P_2 the multiple of 5 must be adjacent to the multiple of 6. In a run of size 8 or more, the unique multiple of 7 must occur in the intersection of the first 7 integers and the last 7; similarly the unique multiple of 6 must occur in the intersection of the first 6 integers and the last 6. Hence the multiple of 6 and multiple of 7 differ by at most 4. By Property 7, there is no run in which the difference is 4. By Property 5, $141^{[8]}$ is the only run of size at least 8 in which the difference is 2 or 3. Hence, in every other run of size 8 or more, the difference must be 1. ∎

PROPERTY 10. *Except for the maximal run* $141^{[8]}$, *in any run of size* 8 *or more in P_2 the multiple of* 6 *is of the form* $6^a 2^{6b}$ *or* $6^a 3^{6b}$, *where* $a \geq 1$ *and* $b \geq 0$ *are integers of opposite parity.*

Proof. Let $R \neq 141^{[8]}$ be a run of size 8 in P_2, and let n be its multiple of 6. In some order, its multiples of 5 and 7 are $n-1$ and $n+1$, by Property 9. Put $n = 2^c 3^d$, where c and d are positive integers.

Suppose $5 \mid n-1$ and $7 \mid n+1$. Then $2^c 3^d \equiv 3^{3c+d} \equiv 1 \pmod 5$ and $2^c 3^d \equiv 3^{2c+d} \equiv -1 \pmod 7$, so $3c + d \equiv 0 \pmod 4$ and $2c + d \equiv 3 \pmod 6$. Hence $c \equiv d \pmod{12}$ and $c \equiv d \equiv 1 \pmod 2$. Put $a = \min\{c, d\}$, where $a \geq 1$ is odd. Also let $|c - d| = 12s$, and put $\max\{c, d\} = a + 6b$, where $b = 2s \geq 0$ is even. Then $n = 6^a 2^{6b}$ or $6^a 3^{6b}$.

Now consider $5 \mid n+1$ and $7 \mid n-1$. In this case $3c + d \equiv 2 \pmod 4$ and $2c + d \equiv 0 \pmod 6$, so $c \equiv d + 6 \pmod{12}$ and $c \equiv d \equiv 0 \pmod 2$. Put $a = \min\{c, d\}$, where $a \geq 2$ is even. Also let $|c - d| = 12s + 6$, and put $\max\{c, d\} = a + 6b$, where $b = 2s + 1 \geq 1$ is odd. Again we have $n = 6^a 2^{6b}$ or $6^a 3^{6b}$. The property follows. ∎

The combined weight of Properties 5 to 10 now enables us to prove

THEOREM 2. *In P_2, up to 10^{25} there is no run of size 9 and the only maximal runs of size 8 are $141^{[8]}$ and $212^{[8]}$.*

Proof. Property 10 provides a strong restriction on the possible multiples of 6 in any run of size 8 or 9 in P_2. Up to 10^{25} there are just 90 of the special multiples of 6 with a odd, and 84 multiples with a even. It is a straightforward computation to check beside these 174 numbers. We find one gem, the maximal run $212^{[8]}$. No later run of size 8 or 9 occurs up to 10^{25}. ∎

Indeed, among runs in P_2 that contain three consecutive integers which are multiples of 5, 6 and 7, the only other instances of size greater than 3 below 10^{25} are $2302^{[7]}$ and $24575^{[5]}$. These computations provide strong evidence for Conjecture 1.

The study of maximal runs in P_2 raises other intriguing questions, including: Are there infinitely many pairs of consecutive integers in P_2? What is the smallest positive integer r_0 such that for each $r \geq r_0$ there are only finitely many runs of size r in P_2? We simply don't know the answers to these questions, just as we don't know the answers to the corresponding questions for P_1.

Next we turn our attention to runs in P_n for $n \geq 3$.

6. Computing runs in P_n for $n \geq 3$

Computation sheds some interesting light on runs in P_n with $n \geq 3$, and turns up some delightful gems.

Since a and $a + 1$ are coprime, if $\{a, a + 1\} \subset P_n$, then each has n principal divisors, so $a(a + 1) \geq p_1 p_2 \cdots p_{2n}$, the product of the first $2n$ primes. But $(a + 1)^2 > a(a + 1)$, so $a + 1 > (p_1 p_2 \cdots p_{2n})^{1/2}$, and the ceiling of this square root is a lower bound for $a + 1$. The product of primes is never a square, so the floor of the square root gives the lower bound

$$a \geq \lfloor (p_1 p_2 \cdots p_{2n})^{1/2} \rfloor.$$

TABLE 2 lists the first maximal run in P_n for $3 \leq n \leq 7$, together with the corresponding factorizations. In each case the first maximal run has size 2. It is interesting to notice how the small primes crowd in and form the majority of principal divisors in the first maximal run of each rank set. It is possible that the first maximal run in P_n always occurs in the interval $[A, 2^n A]$, where $A = \lfloor (p_1 p_2 \cdots p_{2n})^{1/2} \rfloor$. Indeed, each instance in TABLE 2 occurs in the interval $[A, 2^{n/2} A]$.

TABLE 2: First maximal runs $a^{[2]}$

P_n	First maximal run	a	$a+1$
P_3	$230^{[2]}$	2.5.23	3.7.11
P_4	$7314^{[2]}$	2.3.23.53	5.7.11.19
P_5	$254540^{[2]}$	2^2.5.11.13.89	3.7.17.23.31
P_6	$11243154^{[2]}$	2.3.13.17.61.139	5.7.11.19.29.53
P_7	$965009045^{[2]}$	5.7.11.13.23.83.101	2.3.17.29.41.73.109

TABLE 3 lists the starters of successive maximal runs of size 2 in P_n for $3 \leq n \leq 7$. In each case the interval $[A, 2^{n/2}A]$ contains at least 5 maximal runs. Moreover, for P_8, the corresponding interval contains the maximal run $a^{[2]}$ with $a = 68971338435$, since $a = 3.5.17.23.29.31.103.127$ and $a + 1 = 2^2.7.11.13.19.37.107.229$, but we do not know whether this is the first maximal run in P_8, nor how many maximal runs occur in this interval.

TABLE 3: Successive maximal runs of size 2

P_n	Starters of successive maximal runs of size 2
P_3	230, 285, 429, 434, 455, 494, 560, 594, 609, 615, ...
P_4	7314, 8294, 8645, 9009, 10659, 11570, 11780, 11934, 13299, ...
P_5	254540, 310155, 378014, 421134, 432795, 483405, 486590, ...
P_6	11243154, 13516580, 16473170, 16701684, 17348330, 19286805, ...
P_7	965009045, 1068044054, 1168027146, 1177173074, 1209907985, ...

We have also computed the first maximal runs of various sizes $r \geq 3$ in P_n for $3 \leq n \leq 6$. TABLE 4 summarizes this data.

TABLE 4: First maximal runs of increasing size

r	P_3	P_4	P_5	P_6
2	230	7314	254540	11243154
3	644	37960	1042404	323567034
4	1308	134043	21871365	...
5	2664	357642	129963314	
6	6850	2713332	830692265	
7	10280	1217250	4617927894	
8	39693	14273478	...	
9	44360	44939642		
10	48919	76067298		
11	218972	163459742		
12	534078	547163235		
13	2699915	2081479430		
14	526095	2771263512		
15	17233173	...		
16	127890362			

From TABLES 2 and 4 we have $r(3) \geq 16$, $r(4) \geq 14$, $r(5) \geq 7$, $r(6) \geq 3$ and $r(7) \geq 2$. We have also seen that $r(8) \geq 2$. It is noteworthy that the data in TABLE 4 is not monotonic: the first maximal run of size 14 in P_3 precedes the first maximal runs of sizes 12 and 13, and the first maximal run of size 7 in P_4 precedes the first maximal run of size 6. We have already seen this phenomenon in P_2, where the first maximal run of size 8 precedes the first maximal run of size 7.

Let us briefly consider lower bounds for the starters of runs of size $r \geq 3$. Of course, the square root lower bound A for runs of size 2 is also a lower bound for longer runs, but we want something stronger. Suppose $\{a, a+1, a+2\} \subset P_n$. The only divisor that can be common to two of these integers is 2. If $2 \mid a+1$ there is no common divisor, so $a(a+1)(a+2) \geq p_1 p_2 \cdots p_{3n}$, the product of the first $3n$ primes. If $2 \mid a$ then 2 is a common divisor and $8 \mid a(a+2)$, so $a(a+1)(a+2) \geq 4 p_1 p_2 \cdots p_{3n-1}$. Combining $(a+1)^2 > a(a+2)$ with the very weak inequality $p_{3n} > 4$ yields a lower bound that holds regardless of whether a is odd or even:

$$a \geq \lfloor 2(p_2 p_3 \cdots p_{3n-1}))^{1/3} \rfloor.$$

Let us denote this lower bound by B. Similar reasoning yields lower bounds for the starters of longer runs, but here we only consider A and B.

If $n = 2$ then $A = 14$ and $B = 20$, and these are precisely the starters of the first runs of sizes 2 and 3 in P_2. If $n = 3$ then $230^{[2]} \subset [A, 2A]$ and $644^{[3]} \subset [B, 2B]$ where $A = 173$ and $B = 338$. And so on. Perhaps the first maximal run of size 3 in P_n always occurs in the interval $[B, 2^n B]$.

7. Nontrivial runs in P_n for $n \geq 3$

Our computational results certainly confirm that $r(n) \geq 2$ for $n \leq 8$. But it is not obvious that there are nontrivial runs in P_n for every n. In this section we shall show how to make new runs from old, in particular, how to use suitable sets of 4 neighboring integers in P_n to produce pairs of consecutive integers in P_{2n-1}.

For positive integers a, s, t with $s < t$, the integers $a(a+s+t)$ and $(a+s)(a+t)$ differ by st. By imposing appropriate conditions on the divisors of a, $a+s$, $a+t$ and $a+s+t$, we can ensure that

$$\begin{aligned} b &= \frac{1}{st} a(a+s+t) \\ b+1 &= \frac{1}{st}(a+s)(a+t) \end{aligned}$$

are consecutive integers with equally many principal divisors. The simplest case is when $s = 1$ and $t = 2$:

THEOREM 3. *If* $\{a,\ a+1,\ a+2,\ a+3\} \subset P_n$ *and* $12 \mid a(a+3)$, *then* $b = \frac{1}{2}a(a+3)$ *and* $b+1 = \frac{1}{2}(a+1)(a+2)$ *are consecutive integers in* P_{2n-1}.

Proof. Since a and $a+3$ differ by 3, their only possible nontrivial common factor is 3. If $12 \mid a(a+3)$ then exactly one of a and $a+3$ is divisible by 4, the other is odd and both are divisible by 3. Hence the principal divisors of $b = \frac{1}{2}a(a+3)$ include a power of 2, and the product of powers of 3 in a and $a+3$, so $b \in P_{2n-1}$. Also $a+1$ and $a+2$ are relatively prime and $b+1$ is odd, so $b+1 \in P_{2n-1}$. ■

For example, with $a = 33$ and $(s, t) = (1, 2)$, the run $33^{[4]} \subset P_2$ yields $b = 594 = \frac{1}{2}.33.36 = 2.3^3.11$ and $b+1 = 595 = \frac{1}{2}.34.35 = 5.7.17$, so $\{b, b+1\} \subset P_3$. In fact, $594^{[2]}$ is a maximal run in P_3.

By making other choices for a, s, t so that $\{a, a+s, a+t, a+s+t\} \subset P_n^*$, we can construct $\{b, b+1\} = \{a(a+s+t)/st, (a+s)(a+t)/st\} \subset P_m$, where $P_m = P_{2n-1}$ when $P_n^* = P_n$, and $P_m = P_{2n}$ when P_n^* is a suitable union of two or more rank sets. TABLE 5 summarizes some illustrative examples, based on the maximal runs $33^{[4]}$, $141^{[8]}$ and $2302^{[7]}$ in P_2, and $1308^{[4]}$, the first run of size 4 in P_3. Note that extending into the neighborhood of $2302^{[7]}$ yields some pairs $\{b, b+1\}$ in P_4.

Up till this point in our discussions, P_8 is the highest rank set in which we have noted a nontrivial run. With Theorem 3 we can now produce an example in P_9. The first maximal run of size 4 in P_5 is $a^{[4]}$ where $a = 21871365$. Since $a+3$ is a multiple of 12, we have

$$b = \frac{1}{2}a(a+3) = 2^2.3^3.5.29.31.41.137.239.367$$

$$b+1 = \frac{1}{2}(a+1)(a+2) = 7.11.17.23.37.61.97.131.277$$

so $\{b, b+1\} \subset P_9$ with $b = 239178336288660$. Hence $r(9) \geq 2$.

TABLE 5: New runs from old

P_n^*	s	t	a	b	P_m
P_2	1	2	33	594	P_3
P_2	1	2	141	10152	P_3
P_2	1	2	144	10584	P_3
P_2	1	3	141	6815	P_3
P_2	1	3	144	7104	P_3
P_2	1	2	2304	2657664	P_3
P_2	1	3	2303	1771007	P_3
P_2	1	3	2304	1772554	P_3
P_2	1	4	2303	1328831	P_3
$P_2 \cup P_3$	1	3	2300	1766400	P_4
$P_2 \cup P_5$	2	3	2310	891275	P_4
$P_2 \cup P_3$	2	3	2313	893589	P_4
P_3	1	2	1308	857394	P_5

8. Matched primes in arithmetic sequences

In this section we shall show how the simultaneous occurrence of primes in two arithmetic sequences leads to pairs of consecutive integers in P_n. Our approach is based on Dirichlet's famous 1837 theorem on primes in an arithmetic sequence. His presentation can be read in [**3**]; for a more accessible account [**2**] is recommended.

DIRICHLET'S THEOREM ON PRIMES IN AN ARITHMETIC SEQUENCE. *For any coprime positive integers m and r, the arithmetic sequence $\{km + r : k \geq 0\}$ contains infinitely many primes.*

In fact it is known that the primes are shared rather equitably among the arithmetic sequences with common difference (modulus) m. There are $\varphi(m)$ such sequences in which the members are coprime with m, and each contains about $n/\varphi(m)$ of the first n primes that are not divisors of m. As n grows this is increasingly accurate. Indeed, there is a "folk theorem" that among the primes there are infinitely many occurrences of any finite "pattern" that is not explicitly ruled out by modular considerations. For instance, the patterns $\{a, a+1\}$ and $\{a, 2^a + 1\}$ are ruled out modulo 2 and 3 respectively, but $\{a, a+2\}$ and $\{a, 2^a - 1\}$ are "possible" patterns in the primes, and the "folk theorem" applied to them would imply the existence of infinitely many twin primes and infinitely many Mersenne primes. Of course, this does *not* prove these possibilities, but does suggest that they could be true. With such considerations, if we match up the nth terms of any two arithmetic sequences that contain primes, it is plausible that infinitely often the matched pairs are both primes. Formally, this can be stated as follows:

CONJECTURE 2. *If m, m', r, r' are positive integers with $\gcd\{m, r\} = 1$ and $\gcd\{m', r'\} = 1$, and $r \equiv r' \pmod 2$ if $m \equiv m' \pmod 2$, then there are infinitely many positive integers k such that $km + r$ and $km' + r'$ are both prime.*

For example, $2k+1$ and $3k+2$ are simultaneously prime when $k = 1, 3, 5, 9, 15, \ldots$, and the corresponding pairs are $\{2k+1, 3k+2\} = \{3, 5\}, \{7, 11\}, \{11, 17\}, \{19, 29\}, \{31, 47\}, \ldots$. This is of interest in our present context because $b = 3(2k+1)$ and $b+1 = 2(3k+2)$ are consecutive integers, so $\{b, b+1\} \subset P_2$ whenever $2k+1$ and $3k+2$ are primes greater than 3. We deduce that $\{21, 22\}$, $\{33, 34\}$, $\{57, 58\}$, $\{93, 94\}, \ldots$ are pairs of consecutive integers in P_2. Generalizing this example, we have

THEOREM 4. *If Conjecture 2 is true, there are infinitely many pairs of consecutive integers in P_n, for each $n \geq 2$.*

Proof. For some $n \geq 2$ let $m = p_1 \cdots p_{n-1}$ and $m' = p_n \cdots p_{2n-2}$ be the product of the first $n-1$ primes and the next $n-1$ primes, respectively. Let x_0 be the smallest positive solution of the simultaneous congruences $x \equiv 1 \pmod{m}$ and $x \equiv 0 \pmod{m'}$. The Chinese Remainder Theorem (see **[16]**, for example) ensures that x_0 exists, and the general solution is $x \equiv x_0 \pmod{mm'}$. Then there are positive integers r and r' such that $x_0 - 1 = mr'$ and $x_0 = m'r$, so the general solution satisfies

$$x = m'r + kmm' = m'(km + r)$$

$$x - 1 = mr' + kmm' = m(km' + r')$$

where k runs through the integers. Since $\gcd\{x_0, x_0 - 1\} = 1$ we have $\gcd\{mr', m'r\} = 1$, so $\gcd\{m, r\} = \gcd\{m', r'\} = 1$. If Conjecture 2 holds, it follows that there are infinitely many positive integers k such that $p = km + r$ and $q = km' + r'$ are simultaneously prime. If k is large enough, then p and q are not among the first $2n-2$ primes, so $b = mq$ and $b+1 = m'p$ are consecutive integers with n principal divisors (all of which happen to be prime). ■

Although Theorem 4 depends on the unproven Conjecture 2, it is still effective in yielding numerical results, because the proof shows how computation can be used to seek concrete instances of the construction. For example, when $n = 10$ we have

$$m = 2.3.5.7.11.13.17.19.23 = 223092870$$

$$m' = 29.31.37.41.43.47.53.59.61 = 525737919635921.$$

The smallest positive solution to $x \equiv 1 \pmod{m}$ and $x \equiv 0 \pmod{m'}$ is

$$x_0 = 6949903578918639188851,$$

so $x_0 - 1 = mr'$ and $x_0 = m'r$ give $r = 13219331$ and $r' = 31152513206355$. The sequences $\{km + r : k \geq 0\}$ and $\{km' + r' : k \geq 0\}$ are simultaneously prime when $k = 26, 38, 74, \ldots$, so the smallest pair of matched primes is

$$\{p, q\} = \{5813633951, 13700338423740301\}$$

yielding $\{b, b+1\} = \{mq, m'p\} \subset P_{10}$ with $b = 3056447818923499884753870$. Hence $r(10) \geq 2$.

9. Upper bounds on $r(n)$ for $n \geq 3$

In the Introduction we noted that no run of consecutive integers in P_n can contain a multiple of $M = p_1 p_2 \cdots p_{n+1}$, the product of the first $n+1$ primes, so $r(n) < M$. Thus $r(2) < 30$. Eventually we proved $8 \leq r(2) \leq 9$. As a first step toward this result, we showed that no run in P_2 contains more than one multiple of 6, which immediately implies the improved upper bound $r(2) < 12$. We shall now show that the same ideas yield corresponding results for $n \geq 3$.

THEOREM 5. *For any positive integer n, let $N = p_1 p_2 \cdots p_n$ be the product of the first n primes, and let b be the largest integer such that no prime factor of the product $b(b+1)$ exceeds p_n. Then no run of consecutive integers greater than bN in P_n contains more than one multiple of N.*

Proof. For any integer b, if no prime factor of the product $b(b+1)$ exceeds p_n, then $\{b, b+1\} \subset S(p_1, p_2, \ldots, p_n)$ and conversely. By the Neighboring Pairs Theorem, there exists a largest integer b with this property. Then bN and $(b+1)N$ are the last two consecutive multiples of N in $S(p_1, p_2, \ldots, p_n)$. Suppose R is a run of consecutive integers greater than bN in P_n, and suppose R contains at least one multiple of N. Let aN be the smallest multiple of N in R. Then $aN \in P_n$ and $N \in P_n \cap S(p_1, p_2, \ldots, p_n)$, so $a \in S(p_1, p_2, \ldots, p_n)$. But $a > b$, so $a + 1 \notin S(p_1, p_2, \ldots, p_n)$. Thus $a + 1$ contains a prime factor $p > p_n$, and $(a+1)N$ has at least $n + 1$ principal divisors. Thus $(a+1)N \notin P_n$, so $(a+1)N \notin R$, and the theorem follows. ∎

COROLLARY 3. *No run of consecutive integers in P_3 contains more than one multiple of* 30, *and no run of consecutive integers in P_4 contains more than one multiple of* 210.

Proof. Since $N = 30$ is the product of the first 3 primes, and $b = 80$ is the largest integer such that $\{b, b+1\} \subset S(2, 3, 5)$, the theorem ensures that no run of consecutive integers greater than $bN = 2400$ contains more than one multiple of 30. On the other hand, the gap between each pair of consecutive primes up to 2411, the first prime greater than bN, is less than 30, with one exception. The exceptional pair is $\{1327, 1361\}$, with difference 34. Since 1350 is the only multiple of 30 between 1327 and 1361, it follows that among the nonnegative integers up to $(b+1)N = 2430$, there is at least one prime between every pair of consecutive multiples of 30. Hence no run of consecutive integers in P_3 contains two consecutive multiples of 30.

A similar argument applies for P_4, with $N = 210$ and $b = 4374$. The gaps between consecutive primes up to 918563, the first prime greater than bN, are all less than 210, so the claimed result follows. Indeed, the largest gap between consecutive primes up to 918563 is 114, achieved by the pair $\{492113, 492227\}$. ∎

Corollary 3 immediately implies bounds which are better than $r(3) < 2.3.5.7 = 210$ and $r(4) < 2.3.5.7.11 = 2310$, but presumably they are still far from the true values.

COROLLARY 4. *Upper bounds on the size of maximal runs in P_3 and P_4 are $r(3) \leq 59$ and $r(4) \leq 419$.*

10. Closing remarks

An extension of Størmer's Neighboring Pairs Theorem shows that for any finite set of primes P and any positive constant c there are only finitely many pairs of integers in $S(P)$ which differ by c. As noted in [**8**], this follows from a theorem of Alan Baker on logarithms of algebraic numbers [**1**].

One of the most intriguing questions left open in our discussion is whether there are pairs of consecutive integers in P_n for every $n \geq 1$. Examples for $1 \leq n \leq 10$ are found in the paper, but what is the case for larger n? We boldly conjecture that in fact there are infinitely many such pairs for every $n \geq 1$. We showed that this holds for $n \geq 2$ if Conjecture 2 is true. Although Conjecture 2 does not appear to imply that P_1 contains infinitely many pairs of consecutive integers, it does imply that there are infinitely many twin primes, an assertion that is a notorious unproven conjecture in its own right.

Apart from our computational results, we have shed little light on the existence of runs of size 3 in P_n. We found such runs for $1 \leq n \leq 6$, but we have no example with $n \geq 7$, and no basis for conjecturing whether or not any P_n might contain infinitely many such runs. But it is conceivable that for every $n \geq 1$ there are only finitely many runs of size greater than N in P_n, where N is the product of the first n primes. We showed that this is true when $n = 1$, and we gave strong computational evidence in its favor when $n = 2$.

We intend to publish further computational results in a sequel to this article.

Acknowledgements. The first author is grateful for hospitality, during much of the work on this paper, from the School of Mathematical and Physical Sciences of The University of Newcastle, where he holds a conjoint professorship. Both authors wish to thank Dr. Jason Kimberley for independently verifying the computational results presented in the various tables and correcting two procedural errors in a draft version of Table 4.

REFERENCES

1. A. Baker, Linear forms in the logarithms of algebraic numbers (IV), *Mathematika* **15** (1968) 204–216.
2. H. Davenport, *Multiplicative Number Theory*, 2nd ed., 1980, Springer, Berlin.
3. G. L. Dirichlet, *Lectures on Number Theory*, trans. J. Stillwell 1999, Amer. Math. Soc., Providence, RI.
4. E. F. Ecklund, Jr. and R. B. Eggleton, Prime factors of consecutive integers, *Amer. Math. Monthly* **79** (1972) 1082–1089.
5. E. F. Ecklund, Jr. and R. B. Eggleton, A note on consecutive composite integers, this MAGAZINE **48** (1975) 277–281.
6. R. B. Eggleton and W. F. Galvin, Upper bounds on the sum of principal divisors of an integer, this MAGAZINE **74** (2004) 190–200.
7. The 44th Mersenne prime, with exponent $n = 32582657$, was discovered on Sept. 4, 2006. GIMPS: The Great International Mersenne Prime Search. `http://www.mersenne.org/prime.htm`
8. G. D. Halsey and E. Hewitt, More on the superparticular ratios in music, *Amer. Math. Monthly* **79** (1972) 1096–1100.
9. D. H. Lehmer, On a problem of Størmer, *Illinois J. Math.* **8** (1964), 57–79.
10. D. H. Lehmer, The prime factors of consecutive integers, *Amer. Math. Monthly* **72** (1965), no. 2, part II, 19–20.
11. P. Mihailescu, Primary cyclotomic units and a proof of Catalan's conjecture, *J. Reine Angew. Math.* **572** (2004) 167–195.
12. T. Nagell, *Introduction to Number Theory*, 2nd ed. 1964 Chelsea, New York.
13. C. Størmer, Quelques théorèmes sur l'équation de Pell $x^2 - Dy^2 = \pm 1$ et leurs applications, *Skrifter Videnskabs-selskabet (Christiania) I, Mat.-Naturv. Kl.*, no. 2 (1897), 48 pp.
14. C. Størmer, Sur une équation indéterminée, *C. R. Acad. Sci. Paris* **127** (1898) 752–754.
15. C. Størmer, Solution d'un problème curieux qu'on rencontre dans la théorie élémentaire des logarithmes. *Nyt Tidsskrift for Mat. B* **19** (1908) 1–7.
16. C. Vanden Eynden, *Elementary Number Theory*, 2nd ed. 2001 McGraw-Hill, New York, reissued 2006, Waveland Press, Long Grove, IL.

The Shot Made Round (Across) the Table (Maybe)

JOHN H. RILEY, JR.
Bloomsburg University
Bloomsburg, PA 17815
jriley@bloomu.edu

If you are shooting a billiard ball on an oval table, can you make it return to its starting point? If you can, can you make it hit the cushion an arbitrary number of times? If you are constrained to shoot the ball through a given part of the table, can you make it return to the same point? How many ways can you shoot the ball so that it hits the cushion a specified number of times and returns to its starting point? Can you make the ball come back the way you shot it? For example, can you shoot the ball so that its path is a V or a W (starting and ending at the same point)? If one of these shots is possible at one place on the table, can it be made at another place on the table? Of course, which of these "trick shots" are possible depends upon the shape of the table. We limit ourselves to an elliptical table (without pockets) and show that even then the answers change as the ellipse becomes "less circular."

The problem of describing the path of a billiard ball is not new (e.g., [**11**]). There are many interesting aspects of this problem (see [**2**], [**3**], [**7**], [**9**], [**15**] and the more advanced [**1**], [**8**] and [**12**]). Some readers have seen the elaboration of the billiard ball on a rectangular table given by the animated expositor in *Donald in Mathmagic Land* [**18**]. Although a little dated, this film is still recommended viewing. The paths on an elliptical table are animated nicely on the site [**14**]. The path of the ball on a rectangular billiard table is well understood as is the path on a circular table. One would expect that the elliptical table would resemble the circular table. This is true to a limited extent. There are a few shots which can be made everywhere on a circular table but not on an elliptical one. There are some shots on an elliptical table which can't be made at all on a circular table.

Although our discussion and results are primarily phrased geometrically, it is helpful to also have the perspective of discrete dynamical systems. In fact, the billiard ball problem illustrates the subtleties of dynamical systems. Our central question, about when and how a billiard ball returns to its starting point, is really about the existence and nature of periodic orbits for the dynamical system of the billiard ball on an elliptical table. Some of the shots we find can do interesting things before returning to the original point illustrating that periodic orbits in a dynamical system may not be simple.

Notation and terminology

To begin, we show how the path of a billiard ball is a discrete dynamical system. To do this we introduce some notation.

Let X be a space (a set with some structure), let $f : X \to X$ be a function from X to itself and let $f \circ f$ be the composition of f with itself. The *iterates* of f are the functions, one for each positive integer, $f^1 = f$, $f^2 = f \circ f$, $f^3 = f \circ f \circ f, \ldots$. In addition, f^0 is the identity function.

If $x_0 \in X$, the sequence of points x_0, $f(x_0)$, $f^2(x_0)$, $f^3(x_0), \ldots$ is the *orbit* of x_0 under f. If for some $n > 0$, $f^n(x_0) = x_0$, we say x_0 is a *periodic* point of f and the least such n is the *period* of x_0 under f. Periodic points are very important in understanding dynamical systems as they represent the most predictable behavior of the system. (Wait a while and repetition occurs.) If x_0 is n-periodic, the points x_0, $f(x_0)$, $f^2(x_0)$, ... $f^{n-1}(x_0)$ are an n-cycle. (See **[4]**, **[6]**, or **[13]** for basic information about dynamical systems.) In some situations, periodic points also represent the long term behavior of the system in the sense that some nonperiodic orbits may get close to a periodic orbit. This actually happens on an elliptical billiard table—some of the shots will "settle down" after a while (see Theorem 5).

We assume the ellipse is in a coordinate plane with equation $x^2/a^2 + y^2/b^2 = 1$. Sometimes we will consider it to be parameterized by $x = a\cos\theta$, $y = b\sin\theta$, $0 \le \theta < 2\pi$. By assuming $0 < b < a$ the major axis of the ellipse will always be along the x-axis and the foci are at $(\pm c, 0)$ where $b^2 + c^2 = a^2$. Circles are excluded by the condition $b < a$.

How much an ellipse misses being a circle is given by its *eccentricity*, e which is defined by $e = \sqrt{1 - (b^2/a^2)}$. When $e = 0$, the ellipse is a circle and as $e \to 1$, the ellipse becomes flatter.

We are interested in the *path* of a billiard ball in the ellipse. The path may be described by the sequence of points $P^0, P^1, P^2, \ldots$ on the ellipse where it is reflected, with the understanding that between the points the ball travels in a straight line. Being reflected means that if P^0, P^1, P^2 are points on the path where the path is reflected, the line segments P^0P^1 and P^1P^2 make equal angles with the tangent line to the ellipse at P^1. Alternatively, P^0P^1 and P^1P^2 make equal angles with the normal to the ellipse at P^1. See FIGURE 1.

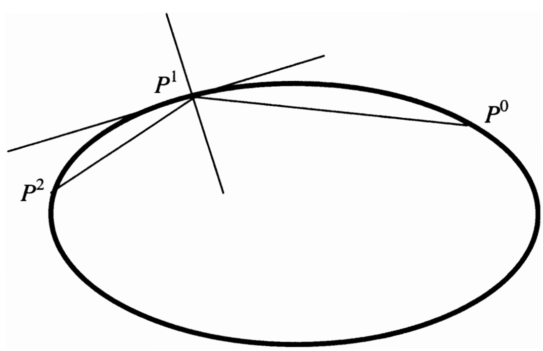


Figure 1 Reflection in an ellipse.

As a matter of convenience, a point on a path will usually refer to a point on the path where the path is reflected. A path includes these points as well the line segments between them. This means all the billiard shots start at the edge of the table and keep going (forever) from points on the edge of the table by bouncing off the cushion.

Notice that the next point on a path depends on the previous two points. If $\mathcal{E}$ denotes the ellipse, to describe the path of the ball we need a function $R : \mathcal{E} \times \mathcal{E} \to \mathcal{E} \times \mathcal{E}$ where $R(P, Q) = (Q, N(P, Q))$ in which P denotes the previous point, Q the current point and $N(P, Q)$ the next point of the path on the ellipse. As a dynamical system, $\mathcal{E} \times \mathcal{E}$ is the underlying space and R is the function of interest. The path can also be specified by the points and angles or directions of reflection. In this case we have $R : \mathcal{E} \times (0, \pi) \to \mathcal{E} \times (0, \pi)$ or $R : \mathcal{E} \times V^2 \to \mathcal{E} \times V^2$ where V^2 is the space of two-dimensional vectors. The map R is continuous. Although R is continuous, changing the starting point by an arbitrarily small amount may change the nature of the periodic orbits that exist. In other words, if I make a trick shot from a point and then challenge you to make the same shot but from a slightly different point, you may not be able to. This isn't due to our limitations as billiard players (after all we're doing this in a mathematical setting so we make perfect shots) but to the table itself.

In the language of dynamical systems a billiard ball's path is an n-cycle if the points on it are an n-cycle using the reflection map R. The preceding indicates that a billiard ball's path is an n-cycle if and only if there are two successive points on the path that replicate two previous points on the path which occurred successively. More precisely, a path is an n-cycle if there is an integer k so that $P^k = P^{k+n}$ and $P^{k+1} = P^{k+1+n}$. If this holds for one k, it holds for all k.

The circle

Our questions about paths in elliptical billiard ball tables are motivated by asking how closely an elliptical table resembles a circular table. When a billiard ball starts in a circular table, all the angles of reflection in a path are the same and the map becomes essentially one-dimensional. This case is easily described as follows. Let α be the angle of reflection in a path (i.e., either of the angles made by the ball and the tangent line to the circle). Then it is not difficult to show that if $\alpha = \pi p/n$ where p/n is a fraction in reduced form, then the path of the ball is an n-periodic orbit. In fact, if n is a positive integer, there is a distinct n-periodic orbit for each integer $p < n/2$ which is relatively prime to n. (We consider two paths to be the same if one is just the other in reverse order.) For example, there are two 5-cycles in a circle: the pentagon (corresponding to $p = 1$) and the pentagram (corresponding to $p = 2$). (See FIGURE 2). So there are two ways to shoot a billiard ball from a cushion on a circular table so that it returns to the same place after hitting the cushion four times.

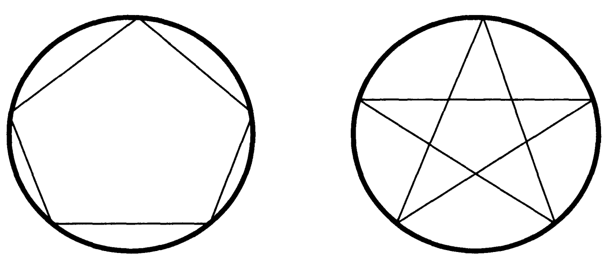


Figure 2 5-cycles in a circle.

Notice if you stand at the center of the circle and turn around to watch the points on the n-periodic orbit corresponding to p, you will turn through an angle of $2\pi p$, i.e., make p rotations. If α is an irrational multiple of π then the path is not periodic and the points where the ball reflects off the circle are dense in the circle.

There is another observation to make about the path of the billiard ball in the circle. Unless the path is a reflection back and forth across a diameter of the circle (i.e., a 2-cycle), all the line segments in the path are tangent to a common interior circle. One could even consider the 2-cycles as tangent to a degenerate circle (the center point).

The preceding description applies to all the points on the circle. Moreover, for each n and each p relatively prime to n, the n-cycles corresponding to p at all the points of the circle are tangent to the same circle.

After some preliminaries, our examination of ellipses begins by showing one difference between a circle and an ellipse.

Preliminaries

It is an exercise (literally—[**10**, p. 663, #110], or [**17**, p. 727, #59]) to prove the reflecting property of the ellipse as illustrated in FIGURE 3. For the sake of reference, we record this as:

PROPOSITION 1. *The normal to an ellipse at a point bisects the angle at the point formed by the line segments from the point to the two foci of the ellipse.*

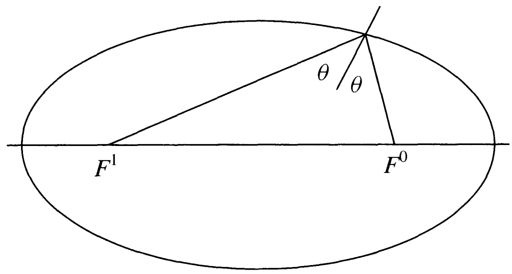


Figure 3 The normal bisects the angle drawn to the foci.

An immediate consequence of Proposition 1 is that if one line segment of a path crosses the major axis between the foci, then all the line segments of the path cross the major axis between the foci. This implies that points on such a path alternate between the upper and lower half of the ellipse.

Using Proposition 1 and the characterization of an ellipse as the loci of points whose sum of distances from the two foci is constant, one can prove the following theorem, which should be better known.

THEOREM 2. *Let F^0 and F^1 denote the foci of an ellipse and F^0F^1 the line segment between the foci.*

(1) *If a path intersects F^0F^1 once but not F^0 or F^1, every line segment in the path is tangent to one of the two branches of a common hyperbola. The foci of the hyperbola are the foci of the original ellipse (i.e., F^0 and F^1).*
(2) *If one line segment in a path does not intersect F^0F^1, every line segment in the path is tangent to a common ellipse. The foci of the ellipse are the foci of the original ellipse.*

Theorem 2 is mentioned in [**16**] and proved in [**1**]. Ambitious readers (with a little bit of time) may use Mathematica® or Java™ to see what Theorem 2 looks like. Equally ambitious readers without time may just go to [**14**]. Part (2) of the theorem shows that some of the paths in an elliptical table resemble those in a circular table. Instead of being tangent to a common circle, the paths in (2) are tangent to a common ellipse.

A priori, it might be possible to shoot a billiard ball back to the same point on the ellipse and have it return at different angle. Theorem 2 implies that if a path returns to the same point on an ellipse, it can't be reflected at a different angle. If the path returns to the same point, the path is actually an n-cycle. In other words, if the ball returns to its starting point on the cushion its path will keep repeating. We state this precisely.

COROLLARY 3. *Suppose $P^0, P^1, \ldots, P^n$ are points on a path with $P^0 = P^n$ (and $P^i \neq P^j$ for $0 \leq i < j < n$). Then the points are on n-cycle, i.e., $P^{n+k} = P^k$ for every integer k.*

Proof. Unless the path is a reflection (i.e., a 2-cycle) along the major axis, it can't contain a focus. (2-cycles are discussed below.) If the path doesn't contain a focus, P^0P^1 and $P^nP^{n+1} = P^0P^{n+1}$ must be tangent to the same ellipse or hyperbola. So either $P^0P^1 = P^nP^{n+1} = P^0P^{n+1}$ or $P^0P^1 = P^nP^{n-1} = P^0P^{n-1}$. Since $P^{n-1} \neq P^1$ we must have $P^{n+1} = P^1$. ■

The reader should determine what would cause the path to have $P^0 = P^n$ and $P^1 = P^{n-1}$.

Elliptical 2-cycles

In contrast to the circle where there are 2-cycles at every point the ellipse has only two 2-cycles. It's only possible to bounce the ball directly back and forth along the axes of the ellipse.

THEOREM 4. *The only 2-cycles in an ellipse are along its axes.*

Proof. It is clear that the paths between $(a, 0)$ and $(-a, 0)$ and between $(0, b)$ and $(0, -b)$ are both 2-cycles since the tangent lines at these pairs of points are parallel.

Suppose P^0, P^1 is a 2-cycle. Then the line containing P^0 and P^1 is perpendicular to the tangents at P^0 and P^1. Let F^0 and F^1 denote the foci of the ellipse. If P^0 and P^1 are not the vertices of the ellipse ($(\pm a, 0)$ in the coordinate plane), then by Proposition 1, triangles $P^0F^0P^1$ and $P^0F^1P^1$ are congruent (angle-side-angle). So the lengths of the line segments P^0F^0 and P^0F^1 are equal. The only points on the ellipse where the distances to the foci are equal are on the minor axis ($(\pm b, 0)$ in the coordinate plane). This shows the path is along the minor axis. ■

The 2-cycle in the ellipse along the major axis (between $(\pm a, 0)$ in the coordinate plane) has a property with no analog in the circle: paths through foci converge to it. If you shoot a billiard ball through a focus and wait long enough the ball will almost be going back and forth between the vertices. In brief, the foci focus. This property is mentioned in **[16]** but not proven there. It is proven in **[5]**, but the proof we give has a slightly different flavor.

THEOREM 5. *Suppose P^0 and P^1 are points on a path $P^0, P^1, P^2, P^3, \ldots$ in an ellipse and a focus is on the line segment P^0P^1. Then the points $P^0, P^2, P^4, \ldots$ converge to one vertex and the points $P^1, P^3, P^5, \ldots$ converge to the other vertex.*

Proof. Because each line segment on the path must pass through a focus the points $P^0, P^2, P^4, \ldots$ are on one side of the major axis and the points $P^1, P^3, P^5, \ldots$ are on the other. Let F^1 be the focus on P^0P^1 and F^0 the other focus. Let V^i be the vertex closest to F^i (see FIGURE 4). For nonnegative integers i, let A^{2i} be the magnitude of the angle $V^0F^0P^{2i}$. Since P^{2i+2} is on the side of the triangle $F^0P^{2i}P^{2i+1}$ extended beyond the triangle, the sequence $\langle A^{2i}\rangle$ is decreasing. As this sequence is bounded below by 0, it must converge to a limit A. Because the map $A^{2i} \to A^{2i+2}$ is the result of two reflections, it is continuous, and A must be a fixed point of two reflections. The only angle that is fixed in this situation is 0, so $A = 0$. This implies $\langle P^{2i}\rangle$ converges to V^0. Correspondingly, $\langle P^{2i+1}\rangle$ converges to V^1. ■

Theorem 4 describes the 2-cycles of an ellipse. Theorem 5 describes the paths through the foci. The rest of the paper is concerned with n-cycles in the ellipse where $n > 2$. These are divided into two categories. The first category consists of the n-cycles which go around the focal segment. These resemble the cycles of the circle. The second category consists of the n-cycles which go through the open focal segment. The

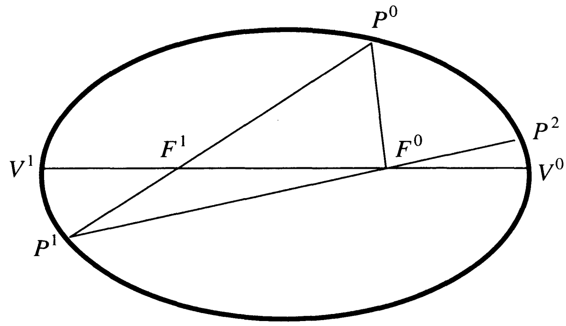


Figure 4 Path through foci.

n-cycles in this category do not have analogs in the circle and are harder to analyze, but correspond to new and more interesting trick shots.

Elliptical n-cycles, $n > 2$, going around the closed focal segment

Theorem 6 gives a complete characterization of cycles that go around the focal segment. It shows that when a circle is stretched into an ellipse only 2-cycles are lost. In other words, on an elliptical table you can make almost all of the trick (returning) shots you could make on a circular table.

THEOREM 6. *Suppose P^0 is a point on an ellipse and n is an integer, $n \geq 3$. Then for every $p < n/2$ which is relatively prime to n, there is a unique n-cycle which includes P^0, and does not intersect the line segment between the foci. These are the only n-cycles through P^0 which do not intersect the line segment between the foci.*

Proof. For the sake of definiteness, assume P^0 is in the first quadrant (including $(0, b)$ but not $(a, 0)$) and let F be the focus on the negative x-axis. Let t_F be the acute angle formed by FP^0 and the tangent to the ellipse at P^0. For each t, $0 < t \leq t_F$ we consider the points P^0, $P^1(t)$, $P^2(t), \ldots, P^n(t)$ on the path which starts at P^0 making an angle t with the tangent at P^0 (i.e., the acute angle formed by $P^0P^1(t)$ and the tangent is t). Intuitively speaking this means we are considering counter-clockwise paths. Using the positive focus would yield clockwise paths and lead to the same cycles, but in the opposite direction. FIGURE 5 illustrates the location of P^0, $P^1(t_F)$, and $P^n(t_F)$

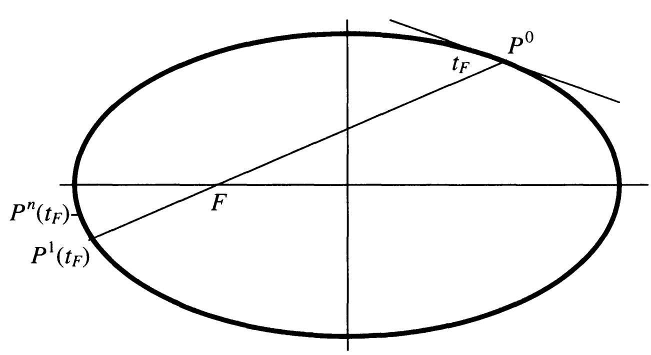


Figure 5 nth point on a path through a focus, n odd.

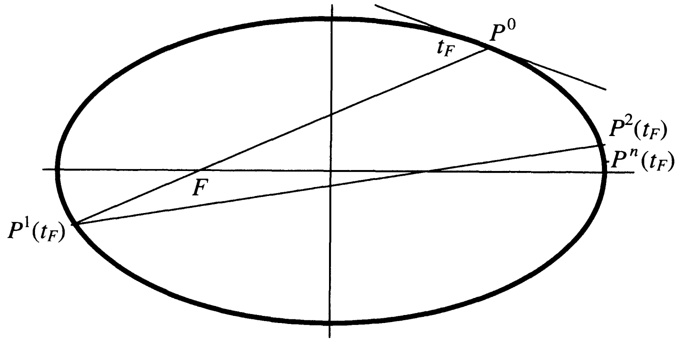


Figure 6 *n*th point on a path through a focus, *n* even.

when n is odd and FIGURE 6 when n is even. The focusing property of the ellipse (Theorem 5) implies that $P^n(t_F)$ is closer to the vertex $(a, 0)$ than P^0 when n is even and $P^n(t_F)$ is closer to the vertex $(-a, 0)$ than $P^1(t_F)$ when n is odd.

Notice that if $0 < t < t_F$ the path goes around the focal segment since $P^0P^1(t)$ does not meet the focal segment and hence by Theorem 2, the entire path goes around the focal segment. There is an n-cycle at P^0 for each t for which $P^n(t) = P^0$ and $P^j(t) \neq P^0$, $1 \leq j < n$. (If $P^j(t) = P^0$ for some $j < n$, there is a cycle, but it is not an n-cycle.)

For each t, $0 < t \neq t_F$ let $\theta^n(t)$ be the counterclockwise rotation around the origin that is made in going from P^0 to $P^1(t)$ to $P^2(t)$ to ... to $P^n(t)$. $P^n(t) = P^0$ if $\theta^n(t)$ is a multiple of 2π. Observe (see FIGURE 5) that if n is odd, $\theta^n(t_F) > 2\pi(n-1)/2 + \pi/2 > 2\pi(n-1)/2$. If n is even, $\theta^n(t_F) > 2\pi(n/2) + 3\pi/2 > 2\pi(n/2)$ (see FIGURE 6). As t decreases from t_F to 0, $P^n(t)$ moves continuously clockwise around the ellipse and $\theta^n(t)$ continuously decreases to 0. Consequently for each integer $p < n/2$, there will be a unique value t_p for which $\theta^n(t_p) = 2\pi p$ and correspondingly $P^n(t_p) = P^0$. Then $P^0, P^1(t_p), \ldots, P^n(t_p)$ will be an n-cycle or include a cycle.

If p is relatively prime to n, the cycle will be an n-cycle as follows. If P^0, $P^1(t_p)$, $\ldots, P^n(t_p)$ is not an n-cycle there is an integer k, $0 < k < n$ with $P^k(t_p) = P^0$. For this to happen, k must divide n and if $\theta^k(t_p) = 2\pi q$ then $2\pi q(n/k) = 2\pi p$. This implies $qn = pk$. Since $q < p$, n and p must have a common factor contradicting p is relatively prime to n.

If p is not relatively prime to n, suppose $n = qr$ and $p = qs$ where r and s are relatively prime and $s < r$. As above, there is an angle t_s such that $P^r(t_s) = P^0$ and $\theta^r(t_s) = 2\pi s$. We claim $t_p = t_s$ which means $P^0, P^1(t_p), \ldots, P^n(t_p)$ is not an n-cycle. If $t_p < t_s$, then the line segments $P^0P^1(t_p)$, $P^t)P^2(t_p), \ldots, P^{n-1}(t_p)P^n(t_p)$ are tangent to an ellipse which contains the ellipse to which

$$P^0P^1(t_s),\ P^1(t_s)P^2(t_s), \ldots, P^{r-1}(t_s)P^r(t_s)$$

are tangent. As a consequence, $\theta^r(t_p) < \theta^r(t_s) = 2\pi s$. By considering tangents again, we see that $\theta^{2r}(t_p) - \theta^r(t_p) < 2\pi s$, $\theta^{3r}(t_p) - \theta^{2r}(t_p) < 2\pi s$ and so on. This would imply $\theta^n(t_p) = \theta^{qr}(t_p) < 2\pi sq = 2\pi p$, a contradiction. The argument that $t_p > t_s$ implies a contradiction is essentially the same (just reverse the inequalities). We conclude $t_p = t_s$ and hence $P^0, P^1(t_p), \ldots, P^n(t_p)$ is not an n-cycle.

A slight modification of the proof yields the result for $P^0 = (a, 0)$. ■

Incidentally, a variation on the proof of Theorem 6 shows that if n is even the sides of the n-cycle occur in parallel pairs. (Start with two points on the ellipse symmetric with respect to the origin and watch what happens as t increases at both points.) So, for example, a 4-cycle (around the focal segment) is a parallelogram. (For the fun of it, show that at least one of these parallelograms is actually a rectangle.)

Theorem 6 shows that, except for 2-cycles, all the n-cycles present in a circle are also present in an ellipse. The theorem also shows that the focal segment in an ellipse is like the center of a circle for these cycles.

Recall that reflection can be regarded as a dynamical system, $R : \mathcal{E} \times (0, \pi) \to \mathcal{E} \times (0, \pi)$ (the reflection map R takes a point, angle pair to a point, angle pair). Theorem 6 asserts that periodic points are dense in $\{(P, t) : P \in \mathcal{E},\ t \in (0, t_{F-}) \cup (t_{F+}, \pi)\}$ ($t_{F\pm}$ are the angles from the tangent to the foci at P as in the proof of Theorem 6).

Elliptical n-cycles, $n > 2$, passing through the open focal segment

This section contains results about n-cycles passing through the open focal segment. These n-cycles are much harder to describe than the n-cycles which do not meet the closed focal segment. If we think of these as trick shots which go through the open focal segment and return to the starting point (after a specified number of bounces), some shots may be entirely impossible, others may be possible on some elliptical tables but not others. In every case, the shots may not be possible at some points. So you must be more careful when you attempt these shots.

For the sake of brevity, in this section "through F^0F^1" means through the open segment between F^0 and F^1.

A first hint that results about cycles through F^0F^1 are not as general as in Theorem 6 comes from considering such cycles at $(\pm a, 0)$. It's clear the only cycle which includes $(a, 0)$ (or $(-a, 0)$) and a point between F^0 and F^1 is the 2-cycle along the major axis. So if $n > 2$, there are no n-cycles including $(a, 0)$ through F^0F^1. So you can't make any of the trick shots described below from the narrow end of the table. This shows that, in general, if there are any n-cycles through F^0F^1 they will not exist at some points of the ellipse.

Theorem 7 is another indication that cycles through F^0F^1 will be more special than those around the closed focal segment. Theorem 7 shows that it is impossible to shoot the ball from a cushion through the open focal segment and return to the starting point after hitting two (or four or ...) intermediate cushions.

THEOREM 7. *If n is odd, there are no n-cycles passing through F^0F^1.*

Proof. Suppose P^0, P^1 and P^2 are points on a path and the line segments P^0P^1 and F^0F^1 intersect. As a consequence of Proposition 1, $\angle F^0P^1P^0 = \angle F^1P^1P^2$. This implies that P^1P^2 and F^0F^1 intersect. This shows that $P^0, P^2, P^4, \ldots$ and $P^1, P^3, P^5, \ldots$ are on the opposite sides of the ellipse. Since the only points on the same side of the ellipse as P^0 have even indices, if the path is an n-cycle n must be even. ∎

Since there are no odd cycles through F^0F^1 and the only 2-cycles are along the major and minor axes, your first hope for a trick shot through the focal segment must correspond to a 4-cycle which is what we investigate next.

The key to working with cycles through F^0F^1 is determining what the normals at points on the ellipse do. For example, typically we expect the normal at a point to intersect the y-axis within the ellipse as in FIGURE 7. However, if the ellipse is "flat enough" at $(0, b)$ it is possible to have a normal at a point on the upper half of the ellipse intersect the y-axis below the ellipse. In this situation, some billiard shots become possible. We begin by listing some information concerning normals which will be used to determine how flat is flat enough.

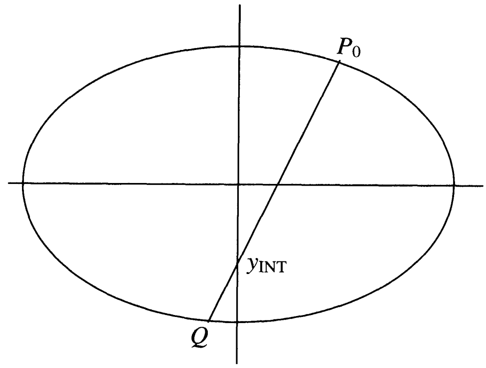


Figure 7 Intersections of a normal line (Proposition 8).

PROPOSITION 8. *Let $P^0 = (x_0, y_0)$ be a point on the ellipse $\mathcal{E}$.*

1. *If P^0 is not $(0, \pm b)$ the normal line at P^0 intersects the y-axis at $y_{\mathrm{INT}} = y_0(1 - a^2/b^2)$.*
2. *The radius of curvature at P^0 is given by*

$$\rho = \frac{(b^4x_0^2 + a^4y_0^2)^{3/2}}{a^4b^4}.$$

3. *If Q is the intersection of $\mathcal{E}$ with the normal line at P^0, then the length of P^0Q is given by*

$$L = \frac{2(b^4x_0^2 + a^4y_0^2)^{3/2}}{b^6x_0^2 + a^6y_0^2}.$$

Proof. All three of these may be derived by considering $\mathcal{E}$ to be parameterized by $\mathbf{r}(\theta) = (a\cos\theta, b\sin\theta)$ and then using standard formulas. (See either **[10]** or **[17]** or virtually any calculus text.) ∎

Since the normal line at $(0, \pm b)$ is the y-axis, strictly speaking, the normal line at $(0, \pm b)$ does not have a y-intercept. However, it is convenient to consider the y-intercept to be given by (1) of Proposition 8. One can consider this to be the limit of the y-intercepts given by (1) as the point P^0 moves to $(0, \pm b)$.

Recall that we are assuming $b < a$. The reader may want to discover the relation between a and b which allows some normals to intersect the y-axis outside the ellipse (above $(0, b)$ or below $(0, -b)$) before reading Proposition 9.

In (3) of Proposition 8, it is useful to call the length of P^0Q (i.e., L) the length of the normal at P^0. In some sense, this is the distance across $\mathcal{E}$ at P^0.

Notice at the vertices $(\pm a, 0)$, $y_{\mathrm{INT}} = 0$ and $\rho = b^2/a < a^2/a < 2a = L$ (because we are assuming $b < a$). Since y_{INT}, ρ, and L are continuous functions of (x_0, y_0) there are intervals of $\mathcal{E}$ containing $(\pm a, 0)$ such that (a) the normal line from every point in the intervals intersects the y-axis between $(-b, 0)$ and $(0, b)$ and (b) the radius of curvature at each point does not exceed the length of the normal at that point. These intervals may or may not be all of $\mathcal{E}$.

Proposition 9 gives a condition which determines whether or not (a) and (b) in the preceding paragraph hold at all points of an ellipse. Ellipses satisfying condition (1) in the proposition are less flat at $(0, \pm b)$ than those that do not, i.e., the ellipses satisfying condition (2). Another way of saying this is that ellipses satisfying condition (1) are closer to being circles. These conditions determine the existence of 4-cycles which cross the focal segment.

PROPOSITION 9. *Suppose the ellipse $\mathcal{E}$ has equation $x^2/a^2 + y^2/b^2 = 1$ with $b < a$.*

(1) *Suppose $2b^2 \geq a^2$ so the eccentricity is $e \leq \sqrt{2}/2$. Then*
   (i) *The normal line from every point of $\mathcal{E}$ intersects the y-axis between $(0, -b)$ and $(0, b)$.*
   (ii) *The radius of curvature at a point of $\mathcal{E}$ does not exceed the length of the normal at that point.*

(2) *Suppose $2b^2 < a^2$ and $e > \sqrt{2}/2$. Then*
   (i) *There are intervals of $\mathcal{E}$ containing $(0, \pm b)$ for which the normal line from every point in the intervals intersects the y-axis outside the segment between $(0, b)$ and $(0, -b)$.*
   (ii) *There are intervals of $\mathcal{E}$ containing $(0, \pm b)$ for which the radius of curvature at each point in the intervals is greater than the length of the normal at that point.*
   (iii) *If the normal at a point intersects the y-axis between $(0, b)$ and $(0, -b)$ then the radius of curvature at the point is less than the length of the normal at that point.*

Proof. (1) Assume $2b^2 \geq a^2$. By the symmetry of the ellipse it suffices to show that (i) and (ii) hold for points in the upper half plane, so we consider a point (x_0, y_0) on $\mathcal{E}$ with $0 \leq y_0 \leq b$.

(i) Since $y_0 \geq 0$, $y_{\text{INT}} = y_0(1 - a^2/b^2) \leq 0$. Also $y_{\text{INT}} = y_0(1 - a^2/b^2) \geq b(1 - a^2/b^2) = (b^2 - a^2)/b \geq (b^2 - 2b^2)/b = -b$.

(ii) Notice $\rho < L$ if and only if $b^6x_0^2 \leq 2a^4b^4$. Since $2b^2 \geq a^2$ and $0 \leq y^0 \leq b$, we find

$$\begin{aligned}
b^6x_0^2 + a^6y_0^2 &= b^4(a^2b^2 - a^2y_0^2) + a^6y_0^2 \\
&= a^2b^6 + a^2(a^4 - b^4)y_0^2 \\
&\leq a^2b^6 + a^2(a^4 - b^4)b^2 \\
&= a^6b^2 \\
&\leq 2a^4b^4
\end{aligned}$$

(2) Assume $2b^2 < a^2$.

(i) Notice that if $y_1 = b^3/(a^2 - b^2)$ then $y_1 = b^3/(a^2 - b^2) < b^3/(2b^2 - b^2) = b$ so there are points on $\mathcal{E}$ with y-coordinate y_1. At such points $y_{\text{INT}} = -b$. Since y_{INT} is a decreasing function of y, for $y_1 < y \leq b$, we have $y_{\text{INT}} < -b$. So for the points on the upper half of $\mathcal{E}$ with $y_1 < y \leq b$, $y_{\text{INT}} < -b$. These points form an interval of $\mathcal{E}$ containing $(0, b)$.

(ii) At $(0, b)$, $\rho = a^2/b > 2b = L$. Since ρ and L are continuous functions of (x, y) this implies there is an interval containing $(0, b)$ on which $\rho > L$.

Finally, symmetry shows there are analogous intervals containing $(0, -b)$.

(iii) Assume that $y_0 > 0$ and that y_{INT} for y_0 is between $(0, b)$ and $(0, -b)$. It follows that $(a^2 - b^2)^2y_0^2 < b^6$. As in the preceding, $\rho < L$ if and only if $b^6x_0^2 + a^6y_0^2 < 2a^4b^4$. Using $(a^2 - b^2)^2y_0^2 < b^6$ we see that

$$
\begin{aligned}
b^6x_0^2 + a^6y_0^2 &= b^4(a^2b^2 - a^2y_0^2) + a^6y_0^2 \\
&= a^2\left(b^6 + \left(\frac{a^2+b^2}{a^2-b^2}\right)(a^2-b^2)^2y_0^2\right) \\
&< a^2\left(b^6 + \left(\frac{a^2+b^2}{a^2-b^2}\right)b^6\right) \\
&= a^2\left(\frac{2a^2b^6}{a^2-b^2}\right) \\
&< a^2\left(\frac{2a^2b^6}{2b^2-b^2}\right) \\
&= 2a^4b^4.
\end{aligned}
$$

It should be noted that in (2) the intervals containing $(0, b)$ for which (i) and (ii) are true are not the same.

Proposition 9 makes it possible to prove that if the ellipse is flat enough, there are places near $(0, b)$ at which 4-cycles through the open focal segment exist. These trick shots are different from the 4-cycles described earlier in Theorem 6 which go around the open focal segment. ■

THEOREM 10. *If $2b^2 < a^2$ and $e > \sqrt{2}/2$ there is an interval of $\mathcal{E}$ containing $(0, b)$ in which every point has a 4-cycle which intersects F^0F^1.*

Proof. If $2b^2 < a^2$, let $x_1 = a^2(a^2 - 2b^2)^{1/2}/(a^2 - b^2)$ and $y_1 = b^3/(a^2 - b^2)$. The normal at (x_1, y_1) intersects $\mathcal{E}$ at $(0, -b)$. If P^+ is a point on $\mathcal{E}$ with coordinates (x_0, y_0) and P^+ is strictly between $(0, b)$ and (x_1, y_1) then the normal at (x_0, y_0) intersects the y-axis below $(0, -b)$. Consequently, the normal at (x_0, y_0) meets $\mathcal{E}$ at a point in the fourth quadrant. A symmetric situation occurs at P^- given by $(-x_0, y_0)$.

Now consider an angle $t \geq 0$ formed in the clockwise direction at the points $P^\pm$ with initial side the normal at each point. Let $Q^\pm(t)$ denote the intersection of the nonnormal sides of the angles with $\mathcal{E}$ corresponding to $P^\pm$. See FIGURE 8 (where the normals are the heavy dashed lines).

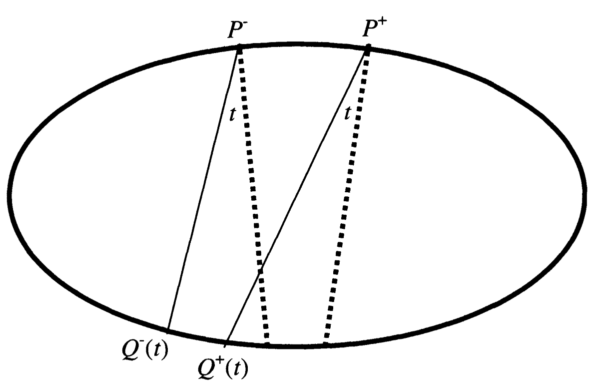


Figure 8 Starting a path at the same angle from symmetric points.

For t close to 0, both of the points $Q^\pm(t)$ are close to the intersection of the normal lines with $\mathcal{E}$. So $Q^+(t)$ starts in the fourth quadrant and $Q^-(t)$ starts in the third quadrant (to the left of $Q^+(t)$). As t increases, both points $Q^{\pm}(t)$ move clockwise. The same value of t, say t_F, causes the side of the angle to pass through the focus on the negative x-axis (consider the angles on the opposite sides of the normals). However, $Q^+(t_F)$ is to the left of $Q^-(t_F)$. By continuity there must be an angle t for which $Q^+(t) = Q^-(t)$. Call this point Q. By symmetry, the same angle in the counterclockwise direction from the normals produces sides which meet at point R. See FIGURE 9. Finally, the symmetry of the situation implies that the line segments P^+R, RP^-, P^-Q and QP^+ are symmetric with respect to branches of the same hyperbola. So the angles at Q and R must be reflections and the four points must form a 4-cycle.

Finally, the normals at $(\pm x_1, -y_1)$ intersect the y-axis at $(0, b)$. So $(0, b)$, $(x_1, -y_1)$, $(0, b)$ and $(-x_1, -y_1)$ form a 4-cycle which includes $(0, b)$. ∎

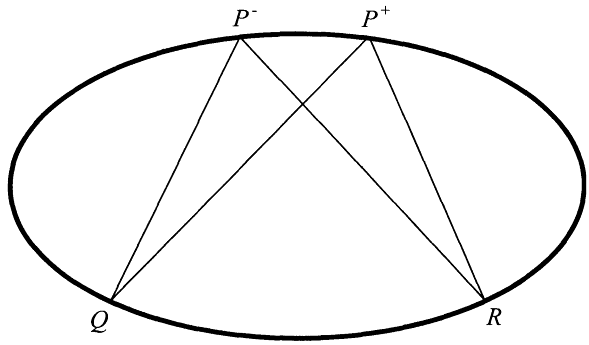


Figure 9 4-cycle through focal segment.

Notice that the 4-cycles of Theorem 10 are symmetric with respect to the y-axis. The extreme example of this is the 4-cycle through $(0, b)$ which is V-shaped. (See the last paragraph of the proof). This would make a rather showy trick shot: shoot the ball from just the right point on the ellipse ($(\pm x_1, -y_1)$ from the proof) to $(0, b)$ and watch it come back on the two lines it went out.

The billiard player who wants to show off with the 4-cycles of Theorem 10 must ensure that the table is flat enough at $(0, b)$. What if $2b^2 < a^2$ isn't true? Can these shots be made? In mathematical parlance, Theorem 10 states $2b^2 < a^2$ is a sufficient condition for 4-cycles through the focal segment; is $2b^2 < a^2$ a necessary condition?

The next results lead to the answer that $2b^2 < a^2$ is necessary for 4-cycles through the focal segment. Lemma 14 is the key.

We recall there is never a 4-cycle through $(\pm a, 0)$ which intersects the focal segment since such a cycle would include the foci and reduce to the 2-cycle along the major axis.

After proving Theorem 10, one would expect that if there are any 4-cycles through the focal segment, there would be one at $(0, b)$ since the ellipse is flattest at $(0, b)$. Lemma 11 shows that $2b^2 < a^2$ is necessary for a 4-cycle through the focal segment at $(0, b)$.

LEMMA 11. *If $2b^2 \geq a^2$ and $e \leq \sqrt{2}/2$, there is no 4-cycle which includes $(0, \pm b)$ and intersects the focal segment.*

Proof. If there is a 4-cycle at $(0, -b)$ which intersects the focal segment, by symmetry it must be in the form of a V which consists of the normals at two points which meet at $(0, \pm b)$. However, if $2b^2 \geq a^2$ normals from points other than $(0, \pm b)$ meet the y-axis strictly between $(0, b)$ and $(0, -b)$ so this cannot occur. ∎

After establishing Lemma 11, it's not surprising that if $2b^2 > a^2$, there are no 4-cycles which intersect the focal segment. To start the proof of this, Lemmas 12 and 13 provide some information about normals.

LEMMA 12. *Let $P(a\cos\theta, b\sin\theta)$ be a point on the ellipse $\mathcal{E}$ and $Q(a\cos T, b\sin T)$ be the intersection of the normal at P with $\mathcal{E}$. If the radius of curvature at P is less than the length of PQ then T is an increasing function of θ (near P).*

Proof. If the radius of curvature at P is less than the length of PQ, the normal is pivoting around a point between P and Q. So if P moves counterclockwise (θ increases) then Q also moves counterclockwise (T increases). See FIGURE 10. ■

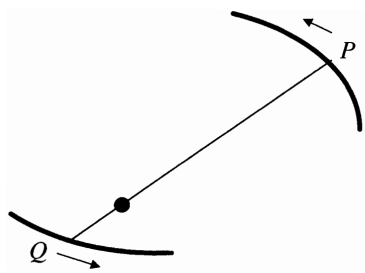


Figure 10 Pivoting around a point in the interior (Lemma 12).

We say two points are in opposite quadrants if either the two points are in the first and third quadrants or the two points are in the second and fourth quadrants. In any ellipse, at each point P there is exactly one point in the opposite quadrant whose normal meets the ellipse at P. (If P is in the first quadrant, and Q moves counterclockwise from $(-a, 0)$ in the third quadrant the intersections of the normals from Q move monotonically from $(a, 0)$ to $(0, b)$ while they are in the first quadrant.) Somewhat nonintuitively, a point on the ellipse (near $(0, \pm b)$) may be the intersection of two normals (from two points on the other side of the ellipse) with the ellipse as in FIGURE 11.

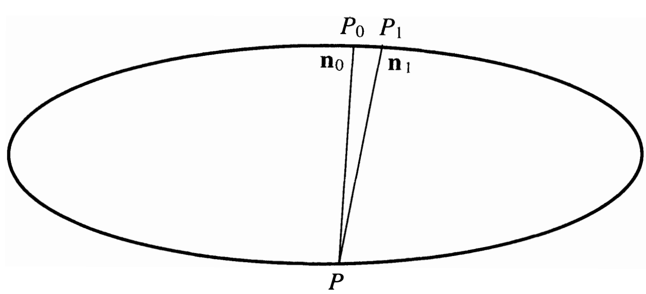


Figure 11 Two normals to the same point.

LEMMA 13. *Suppose the normal from P meets the y-axis between $(0, b)$ and $(0, -b)$. Then there is only one point (which is in the opposite quadrant) whose normal meets the ellipse at P.*

Proof. Suppose P is in the first quadrant and the y-intercept of the normal from P is between $(0, b)$ and $(0, -b)$. As noted above, there is a point in the third quadrant whose normal meets the ellipse at P. If there were a second point Q whose normal met the ellipse at P, Q would be in the fourth quadrant. Since the slope of normals through points in the fourth quadrant are negative, Q would be to the right of P and the normal from Q would meet the y-axis above $(0, b)$. On the other hand, if Q is to the right of P and the normal from P meets the y-axis between $(0, \pm b)$ then the normal from Q meets the y-axis between $(0, \pm b)$, a contradiction. ■

Lemma 14 gives a geometric condition for a point to have no 4-cycle passing through the focal segment. The basic idea of the proof is to start at a point P^0 satisfying the conditions of the lemma and for each angle ϕ with the normal at P^0 consider the points $P^1(\phi)$ and $P^2(\phi)$ which start a path on one side of the normal and $P^{-1}(\phi)$ and $P^{-2}(\phi)$ starting on the other side. By showing $P^2(\phi)$ never equals $P^{-2}(\phi)$ we show no 4-cycle exists. Doing this is tedious, so the proof is omitted.

LEMMA 14. *Suppose P^1 is a point on the ellipse $\mathcal{E}$ whose normal meets the y-axis between $(0, b)$ and $(0, -b)$. Also suppose P^0 is the point in the opposite quadrant from P^1 for which the normal at P^0 meets $\mathcal{E}$ at P^1. Then there is no 4-cycle at P^0 which passes through the focal segment.*

Earlier we noted that at $(a, 0)$ there are no n-cycles through the focal segment (unless $n = 2$). Lemma 14 allows us to extend this to an interval around $(a, 0)$ for 4-cycles. Basically around $(a, 0)$ the ellipse is never flat enough (recall we are assuming $b < a$) to pull off the trick shot of Theorem 10. This means if you are betting you can make a shot of this type, you can't let your opponent choose the starting point.

THEOREM 15. *In any ellipse $\mathcal{E}$ there is a segment of $\mathcal{E}$ containing $(a, 0)$ in which no point has a 4-cycle which intersects the focal segment.*

Proof. At $(-a, 0)$, $y_{\text{INT}} = 0$, so there is an interval I of $\mathcal{E}$ containing $(-a, 0)$ on which normals lie between $(0, b)$ and $(0, -b)$ giving points P^1 as in Lemma 14. There is an interval J of $\mathcal{E}$ containing $(a, 0)$ on which normals meet $\mathcal{E}$ in I giving points P^0 as in Lemma 14. By Lemma 14, J has the desired property. ■

Of course, symmetry gives an interval containing $(-a, 0)$ with the same property.

THEOREM 16. *There exist 4-cycles intersecting the focal segment of $\mathcal{E}$ if and only if $e > \sqrt{2}/2$.*

Proof. If $2b^2 < a^2$, there exist 4-cycles by Theorem 10.
If $2b^2 > a^2$, all normals (except at $(0, \pm b)$) intersect the y-axis between $(0, b)$ and $(0, -b)$. By Lemmas 11 and 14, there are no 4-cycles which meet the focal segment. ■

Ellipses with $2b^2 < a^2$ are flatter (or more elongated) than ellipses with $2b^2 \geq a^2$. Another way of putting is this is that ellipses with $2b^2 < a^2$ are less like circles than ellipses with $2b^2 \geq a^2$. However the preceding theorems suggest that ellipses with $2b^2 < a^2$ are distinctly less like circles in the following sense. If $2b^2 < a^2$ then there are points on the ellipse (near $(0, \pm b)$) with two distinct 4-cycles, one around the focal segment and one through the focal segment. In a circle (or an ellipse with $2b^2 \geq a^2$) every point has only one 4-cycle. You must check your table carefully before attempting to make (or betting you will make) one of these shots.

Having described the situation for 4-cycles through the focal segment we begin investigating 6-cycles through the focal segment. If the table is shaped correctly, more trick shots are possible.

LEMMA 17. *If* $4b^2 < 3a^2$ *(so* $e > \sqrt{1}/2 = 1/2$*) there is a point* (x_1, y_1) *with* $-b < y_1 < 0$ *so that the line from* $(0, b)$ *to* (x_1, y_1) *is reflected through the origin.*

Proof. Let $m_0 = (y_1 - b)/x_1$ be the slope of the line L_0 from $(0, b)$ to (x_1, y_1) and $m_N = a^2y^1/b^2x_1$ be the slope of the normal line at (x_1, y_1). Let L_R be the line which is the reflection of L_0 at (x_1, y_1). (See FIGURE 12.) Then the slope of L_R is given by

$$\frac{-m_0 + m_N^2 m_0 + 2m_N}{1 - m_N^2 + 2m_N m_0}.$$

For L_R to pass through the origin this slope must equal y_1/x_1. Recalling $b^2x^2 = a^2b^2 - a^2y^2$, it is possible to solve the equation for y_1 and obtain $y_1 = b\left(1 \pm \frac{a}{\sqrt{a^2-b^2}}\right)$. The first of these values (using addition) is larger than b and hence doesn't give a point on the ellipse. The second value, $b\left(1 - \frac{a}{\sqrt{a^2-b^2}}\right)$, is negative. Hence for y_1 to be on the ellipse, it suffices to have this value greater than $-b$. From $-b < b\left(1 - \frac{a}{a^2-b^2}\right)$ we obtain $4b^2 < 3a^2$. ■

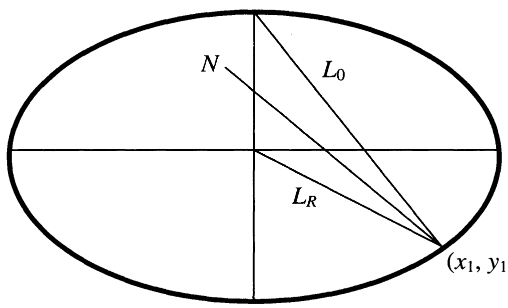


Figure 12 Starting a path at $(0, b)$ which passes through the origin.

THEOREM 18. *If* $4b^2 < 3a^2$ *(hence* $e > 1/2$*) there is a 6-cycle through* $(0, b)$ *which includes the origin (and hence meets the focal segment).*

Proof. If the hypotheses of the theorem are met, by Lemma 17 there is a point (x_1, y_1) so that the path that starts at $(0, b)$ and is reflected at (x_1, y_1) passes through the origin. Since the reflection passes through the origin, it meets the ellipse at $(-x_1, -y_1)$. Symmetry then shows that $(0, b)$, (x_1, y_1), $(-x_1, -y_1)$, $(0, -b)$, $(x_1, -y_1)$, $(-x_1, y_1)$ is a 6-cycle. Since it passes through the origin, it passes through the focal segment. ■

If one follows the points in Theorem 18, one obtains a starlike figure (with six segments) which would be a pretty impressive shot. Nothing like this exists for circles, in which the only 6-cycles are hexagons. These trick shots are only possible on an elliptical table. Of course, Theorem 18 shows there are 6-cycles through the focal segment at points other than $(0, \pm b)$: the points $(\pm x_1, \pm y_1)$. The next theorem shows there is an interval around $(0, b)$ with such cycles. Basically these are obtained by perturbing the 6-cycle through $(0, b)$.

THEOREM 19. *If* $4b^2 < 3a^2$ *and* $e > 1/2$*, there is a segment of the ellipse containing* $(0, b)$ *in which every point has a starlike 6-cycle through the focal segment.*

Proof. Let P^0 be a point on the ellipse in the first quadrant. Let P_F^2 be the point on the ellipse (in the second quadrant) obtained by starting a path from P^0 through the focus F^0 and then going through the focus F^1. Let P_F^{-1} be the point on the ellipse (in the third quadrant) obtained from starting at P^0 and going through F^1. If P^0 is close enough to $(0, b)$, P_F^{-1} will be to the right of (have greater x-coordinate than) P_F^2. (See FIGURE 13, solid lines.)

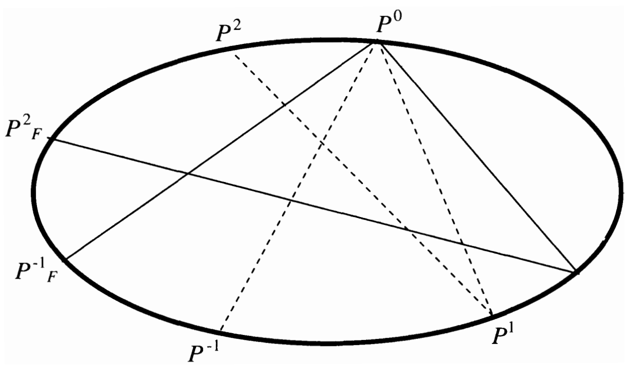


Figure 13 Leading to a 6-cycle (Theorem 19).

Let P^1 be the point on the ellipse (in the fourth quadrant) so that the line through the origin and P^1 is reflected at P^1 to P^0. (P^1 exists by perturbing the path of Theorem 18, keeping it passing through the origin.) Let P^2 be the point on the ellipse from P^1 through the origin and P^{-1} the point on the reflection of P^0P^1 at P^0. (See FIGURE 13, dashed lines.) Since the (absolute value of the) slope of P^0P^1 is greater than the slope of P^0P^{-1} and both lines are tangent to the same hyperbola, P^{-1} is further from the y-axis than P^1. Since P^2 and P^1 are symmetric across the origin, this implies P^{-1} is to the left of P^2.

Let Q^0 be the point obtained by reflecting P^0 across the x-axis and Q^{-1} and Q_F^{-1} be the corresponding points to P^{-1} and P_F^{-1}. (See FIGURE 14.) As before, if we consider what happens as the angle between the normal at P^0 and P^1 increases (and correspondingly the same angle at Q^0 and Q^{-1}) there must be an angle at which P^2 equals Q^{-1}. Then $P^0P^1P^2Q^0$ is the start of a starlike six-cycle which is completed by symmetry. ■

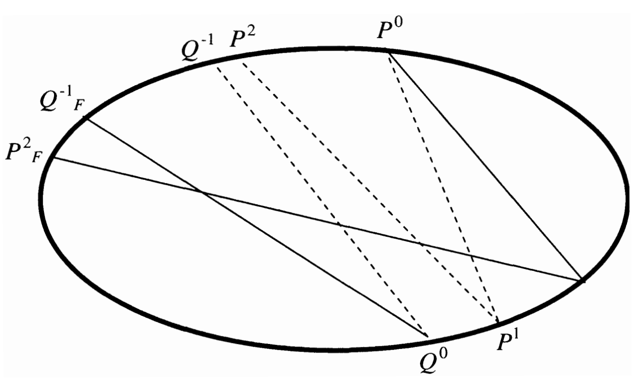


Figure 14 Building a 6-cycle through P^0 and Q^0.

Implicit in the proof of Theorem 19 is that there are also intervals around the points $(\pm x_1, \pm y_1)$ of Theorem 18 with starlike six-cycles, since the original P^1 (of Theorem 19) is (x_1, y_1) perturbed. But the proof does not show the interval at $(0, b)$ extends to these other points although that is the natural suspicion.

Theorem 19 shows that while a circle only has one six-cycle (the hexagon) at a point, some ellipses have two six-cycles at some points. Theorem 20 shows that by flattening the ellipse even more, we obtain an ellipse with three six-cycles at some points. We call these additional six-cycles W-like since they are made of two W's. FIGURE 15 illustrates the three kinds of six-cycles that may occur in an ellipse. The starlike six-cycle is in the upper right and the W-like one in the bottom middle. This means there may be three ways to shoot the ball from $(0, b)$ and have it return after bouncing off five intermediate cushions. If you do all three shots, save the starlike one for last as it's the most impressive (in our opinion).

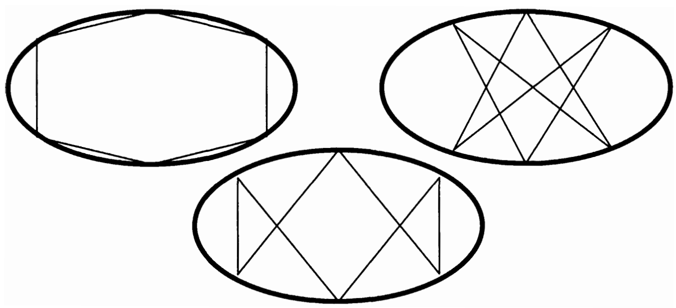


Figure 15 Three different kinds of 6-cycles.

THEOREM 20. *If $2b < a$ there is a segment of the ellipse containing $(0, b)$ in which every point has a W-like 6-cycle through F^0F^1.*

Proof. As before, we assume the point $P^0(x_0, y_0)$ is in the first quadrant. Let l be the line through P^0 which is the reflection of the vertical line at P^0. Then an equation for l is

$$y - y_0 = \left(\frac{a^4y_0^2 - b^4x_0^2}{2a^2b^2x_0y_0}\right)(x - x_0).$$

From this equation we obtain the y-intercept of l:

$$y_{\text{INT}} = \frac{b^4 + (b^2 - a^2)y_0^2}{2b^2y_0}.$$

If $2b < a$, then $b^2/(a - b) < b$.

If $y_0 = b^2/(a - b)$ then $y_{\text{INT}} = -b$. This implies that (x_0, y_0), $(0, -b)$, $(-x_0, y_0)$, $(-x_0, -y_0)$, $(0, b)$, $(x_0, -y_0)$ form a 6-cycle through (x_0, y_0) and $(0, b)$.

Now suppose $b^2/(a - b) < y_0 < b$. Then $y_{\text{INT}} < -b$. Let ϕ be the angle made by the normal and either segment on a path through P^0. We start with $\phi = \phi_V$, the angle made by the normal and the vertical line through P_0. Let $P^1(\phi)$ and $P^2(\phi)$ be the first two points on the path starting at P^0 with angle ϕ on the side of the vertical line at P^0 and $P^{-1}(\phi)$ be the first point in the other direction. The right half of FIGURE 16 illustrates this for $\phi = \phi_V$.

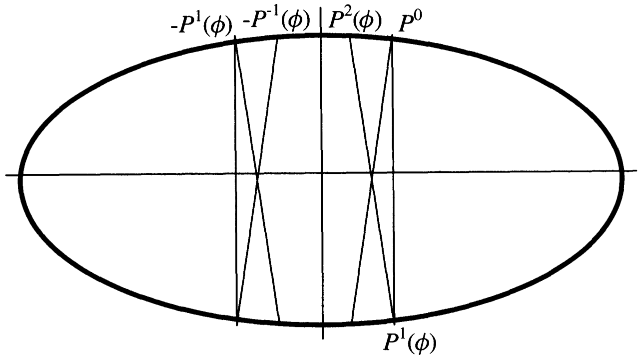


Figure 16 Leading to a W 6-cycle (Theorem 20).

Notice for $\phi = \phi_V$, P^0 and $P^1(\phi_V)$ are symmetric across the x-axis, as are $P^{-1}(\phi_V)$ and $P^2(\phi_V)$. Since $y_{\text{INT}} < -b$, $P^{-1}(\phi_V)$ is to the right of the y-axis and hence so is $P^2(\phi_V)$.

Let $-P^0$ be symmetric to P^0 across the origin. Let $-P^1(\phi)$, $-P^2(\phi)$ and $P^{-1}(\phi)$ be the corresponding symmetric points to $P^1(\phi)$, $P^2(\phi)$ and $P^{-1}(\phi)$.

As ϕ increases from ϕ_V, $P^2(\phi)$ eventually moves counterclockwise (consider what happens when $\phi = \phi_F$) and $-P^{-1}(\phi)$ eventually moves clockwise. By continuity, there will be an angle ϕ_0 for which $P^2(\phi_0) = -P^{-1}(\phi_0)$. By symmetry $-P^2(\phi_0) = P^{-1}(\phi_0)$. Since all the line segments are tangent to the same hyperbola, there are reflections at $P^2(\phi_0) = -P^{-1}(\phi_0)$ and $-P^2(\phi_0) = P^{-1}(\phi_0)$. So the six segments form a 6-cycle through P_0. ■

Conclusions and questions

We have seen that as an ellipse gets flatter, additional cycles (through the focal segment) appear. More trick shots are possible with flatter ellipses. We have given the condition for flatness at (0, b) using a and b since they are used in the proofs. A measure for the "flatness" of ellipses is eccentricity e, $e = \sqrt{1 - (b^2/a^2)}$. Our results, in terms of eccentricity, can be summarized as follows.

Cycles through focal segment

$e > \sqrt{1/2}$	Starlike 6-cycles
$e > \sqrt{2/2}$	4-cycles
$e > \sqrt{3/2}$	W-like 6-cycles

The appearance of cycles in a dynamical system at particular values of a parameter is not unusual (cf. the logistic map **[4]**, **[6]**, **[13]**). The particular values of e in the above are tantalizing. Furthermore, it is not too hard to show that starlike 6-cycles at (0, b) imply starlike 8-cycles at (0, b), so the eccentricity necessary for starlike 8-cycles is smaller than 1/2. Also W-like 8-cycles at (0, b) imply W-like 6-cycles at (0, b) so the eccentricity needed for W-like 8-cycles is greater than $\sqrt{3/2}$. This means the preceding table should be extended. Finding the eccentricity for either class of 8-cycles would be illuminating.

In fact, as the eccentricity increases to 1, more and more W-like (or perhaps more aptly zigzag) higher order cycles (through the focal segment) emerge as follows. As e approaches 1, the ellipse flattens out and, in part, is close to two parallel line segments. Zigzagging between these and using a continuity argument produces the desired cycles. Again, the values of e for which this occurs would be interesting.

Another avenue for investigation is what other kinds of n-cycles through the focal segment are possible. At $(0, b)$, the symmetry of such cycles and the requirement that points are on opposite sides of the ellipse imply only starlike and W-like cycles are possible for 6 and 8-cycles. If $n > 8$ (and even) are there n-cycles (through the focal segment) which are not starlike or W-like? The proofs for 4-cycles and 6-cycles show these are symmetric (if (x, y) is on the cycle, either $(x, -y)$ or $(x, -y)$ or $(-x, -y)$ is on the cycle). Are there asymmetric cycles through the focal segment?

REFERENCES

1. Arno Berger, *Chaos and Chance: An Introduction to Stochastic Aspects of Dynamics*, Walter de Gruter, Berlin, 2001.
2. Duane W. DeTemple and Jack M. Robertson, A billiard path characterization of regular polygons, this MAGAZINE **54** (1981).
3. Duane W. DeTemple and Jack M. Robertson, Convex curves with periodic billiard polygons, this MAGAZINE **58** (1985).
4. Robert L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd ed., Addison-Wesley, Reading, MA, 1989.
5. Marc Frantz, A focusing property of the ellipse, *American Mathematical Monthly* **101** (1994).
6. Denny Gulick, *Encounters with Chaos*, McGraw-Hill, New York, 1992.
7. Eugene Gutkin, Two applications of calculus to triangular billiards, *American Mathematical Monthly* **104** (1997).
8. Benjamin Halpern, Strange billiard tables, *Transactions of the American Mathematical Society* **232** (1977).
9. Charles H. Jepsen, Billiard balls and a number theory result, *Two-Year College Mathematics Journal* **10** (1979).
10. Ron Larson, Robert P. Hostetler, and Bruce H. Edwards, *Calculus with Analytic Geometry*, 7th ed., Houghton Mifflin, Boston, 2002.
11. N. J. Lennes, On the motion of a ball on a billiard table, *American Mathematical Monthly* **12** (1905).
12. Valery Kovachev, Smoothness of the billiard ball map for strictly convex domains near the boundary, *Proceedings of the American Mathematical Society* **103** (1988).
13. Heinz-Otto Peitgen, Hartmut Jürgens, and Dietmar Saupe, *Chaos and Fractals: New Frontiers of Science*, Springer Verlag, New York, 1992.
14. Herman Serras, Elliptical billiard tables, `http://cage.ugent.be/~hs/billiards/billiards.html`
15. Robert Sine and Vladislav Kreinovic, Remarks on billiards, *American Mathematical Monthly* **86** (1979).
16. Hugo Steinhaus, *Mathematical Snapshots*, 3rd American ed., Oxford University Press, Oxford, 1969.
17. James Stewart, *Calculus*, 5th ed., Brooks/Cole, Belmont, CA, 2003.
18. Walt Disney Educational Media Company, *Donald in Mathmagic Land*, Burbank, CA, 1959, 1988.

Erratum

The cover credit on the inside front cover of the June 2008 issue of this MAGAZINE incorrectly identifies Melissa Loe's co-author as Jenny Shepard. She is Jenny Borovsky. (The names are correct on the title page of their article.) We apologize for this error.

Probabilistic Reasoning Is Not Logical

RUMA FALK
The Hebrew University of Jerusalem
Jerusalem, Israel
rfalk@cc.huji.ac.il

Introduction

Consider three uncertain events A, B, and C. We know that when A occurs, C becomes more probable; likewise, the occurrence of B makes C more probable. Now, suppose we have learned that both A and B have occurred. Straightforward common sense probably tells most of us that C has now a fortiori been rendered more probable; or does it?

The ground rules of logic are the stock-in-trade of sensible people who are engaged in reasoning—college students included. Employing principles such as transitivity—that is to say, if event (or proposition) A entails B, and B entails C, then A entails C—is virtually second nature to reasoners. However, judgments are mostly made in uncertain situations. Therefore, a scrutiny of the rules that govern inference under uncertainty is imperative. Tempting as it may be to apply the syllogisms of logical deduction to probabilistic inference, one should be cautious of inadvertently relying on an analogy that might be misleading. The purpose of this article is to examine meaningful rules of inference and establish whether they agree or not when applied to these two modes of reasoning. Recognizing the cases of agreement and disagreement is important for teaching and for scientific inference.

Definitions of relations

In everyday language, words such as imply, involve, entail, necessitate, and the phrase if-then are virtually synonymous. Copi [**5**] repeatedly cautioned that, in formal discourse, one should distinguish between the different senses of "if-then" and "implication" to avoid misunderstandings that may arise from treating a technical term as if it were a term of ordinary language. Hence, to clarify the exposition, the entailment terms will be used here as follows:

Logical Involvement (LI). *A involves B*, denoted $A \rightarrow B$, signifies a conditional proposition to be read "if A then B," where A is the antecedent and B is the consequent. It means that given A, B is a certainty.

Since any event is a subset of the universe, relations between events can be translated in one-to-one correspondence to relations between sets. $A \rightarrow B$ is interpreted in set-theoretic terms as "all As are Bs" or as "A is a subset of B." The LI symbol "$\rightarrow$" denotes an *if-then* relation *within* a proposition, whereas "$\Rightarrow$" will denote an *implication* relation *between* propositions. $X \Rightarrow Y$ means that proposition X implies proposition Y.

Two analogous positive relations between events (sets) come to mind when uncertainty is involved. Knowledge of the occurrence of A might, first, increase the probability of B, second, enhance the probability of B up to near certainty.

Probabilistic Support (PS). *A supports B*, denoted $A \nearrow B$, means that given A, the probability of B increases, that is, $P(B \mid A) > P(B)$. Conversely, $A \searrow B$ means $P(B \mid A) < P(B)$.

Probabilistic Confirmation (PC). *A confirms B*, denoted $A \rightsquigarrow B$, means that given A, the probability of B becomes high, say, greater than 90%, that is $P(B \mid A) > 0.90$. Thus $\rightsquigarrow$ signifies "almost involvement." The threshold 0.90 is arbitrary. Other high probabilities may be chosen without affecting the analysis of the rules that govern the relations.

Rules Governing Relations Between Events

Although it is obvious that if $A \to C$ and $B \to C$ then $A \cap B \to C$, the answer to the question of the validity of that assertion when the relation $\to$ (LI) is replaced by either $\nearrow$ (PS) or $\rightsquigarrow$ (PC)—as posed in the opening paragraph—is not self evident. Likewise, one easily realizes that $A \to B$ implies neither $B \to A$ nor $\overline{A} \to \overline{B}$, and upon a moment's reflection it becomes clear that $A \to B$ and $\overline{B} \to \overline{A}$ are equivalent [**5**, p. 138] (think of A as "it rains" and of B as "it is cloudy"). Yet, it is not immediately clear whether replacing LI by either PS or PC in these cases would yield valid or invalid arguments.

Six rules of inference are listed in the rows of TABLE 1. The rules concern a generic *Relation*, denoted R, between events. R should be replaced with LI, PS, or PC, according to the column's head. "T" in the cell defined by each combination means that the row's rule is *true* when applied to the column's relation, and "F" means that it is *false*. For example, $A \to B \Rightarrow \overline{A} \to \overline{B}$ is false, whereas $A \nearrow B \Rightarrow \overline{A} \nearrow \overline{B}$ is true.

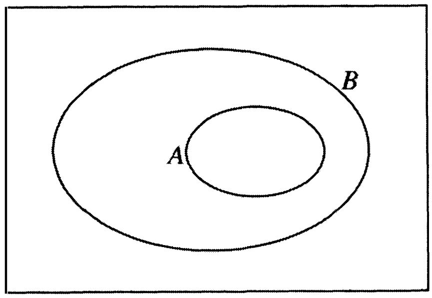
TABLE 1: Truth values of six rules of inference, according to the relation (R) to which they apply (T = True; F = False)

		Relation between events		
	Rule of inference	Logical involvement (LI) $R = \to$	Probabilistic support (PS) $R = \nearrow$	Probabilistic confirmation (PC) $R = \rightsquigarrow$
1	$A\,R\,B \Rightarrow B\,R\,A$	F	T	F
2	$A\,R\,B \Rightarrow \overline{A}R\overline{B}$	F	T	F
3	$A\,R\,B \Rightarrow \overline{B}R\overline{A}$	T	T	F
4	$A\,R\,B\;\&\;B\,R\,C \Rightarrow A\,R\,C$	T	F	F
5	$A\,R\,C\;\&\;B\,R\,C \Rightarrow (A \cap B)RC$	T	F	F
6	$A\,R\,C\;\&\;B\,R\,C \Rightarrow (A \cup B)RC$	T	F	F

The truth values in the 18 cells of TABLE 1 are justified briefly, or heuristically, below, in a semisystematic order. T requires a proof, whereas one counterexample suffices for refuting a general rule in the case of F. To expedite the reading, each case is identified by the cell's coordinates and so are figures or tables that serve to explain that case. I'll dwell particularly on the discussion of cases 3PC and 5PS, which in my experience are not easy for students to accept and which I deem particularly instructive.

Justifications

Cases 1LI, 2LI, and 3LI. The given LI relation, $A \to B$, is represented in FIGURE 1 by $A \subset B$ (A is a proper subset of B), and the three truth values are immediately verified.



$A \subset B; \overline{B} \subset \overline{A}$

Figure 1 $A \to B$ does not imply $B \to A$, likewise, it does not imply $\overline{A} \to \overline{B}$, but it does imply $\overline{B} \to \overline{A}$.

Cases 1PS, 2PS, and 3PS. Assuming $0 < P(A) < 1$, $0 < P(B) < 1$, and given $A \nearrow B$, we know that $P(B \mid A) > P(B)$, that is, $P(A \cap B)/P(A) > P(B)$. Hence $P(B \mid A) > P(B)$, $P(A \cap B) > P(A)P(B)$, and $P(A \mid B) > P(A)$ are all equivalent. This proves case 1PS: $A \nearrow B \Rightarrow B \nearrow A$. Unlike logical involvement, probabilistic support ($\nearrow$) is symmetric, and so is the opposite relation $\searrow$: simply replace $>$ with $<$ in the above derivation. Replacing $>$ with $=$ shows that independence is symmetric [**8**, Problem 2.3.15].

Given $P(B \mid A) > P(B)$, it follows that $P(\overline{B} \mid A) < P(\overline{B})$, namely $A \searrow \overline{B}$, by symmetry, also $\overline{B} \searrow A$, that is, $P(A \mid \overline{B}) < P(A)$. The complementary probabilities of both sides satisfy $P(\overline{A} \mid \overline{B}) > P(\overline{A})$, which means that $\overline{B} \nearrow \overline{A}$. This justifies the T in 3PS. Applying symmetry again, we get $A \nearrow B \Rightarrow \overline{A} \nearrow \overline{B}$, proving the T in 2PS. The didactic lesson to extract from the truth of these three rules concerning PS (see also [**4**] and [**11**]) is that correlation is a symmetric relation and that when A and B are positively (negatively) related so too are $\overline{A}$ and $\overline{B}$.

Case 1PC. A macabre example, credited to Carver [**2**], conclusively *refutes* the notion that $A \rightsquigarrow B$ implies $B \rightsquigarrow A$: Let A be the event that a person was hanged and B the event that the person is dead. Undoubtedly, $P(B \mid A) > 0.90$ and $P(A \mid B) < 0.90$, hence $A \rightsquigarrow B \not\Rightarrow B \rightsquigarrow A$.

Case 2PC. The numerical example in TABLE 2, showing the simultaneous classification of 120 cases according to two attributes, demonstrates that $A \rightsquigarrow B \not\Rightarrow \overline{A} \rightsquigarrow \overline{B}$, because $P(B \mid A) = 0.95 > 0.90$, whereas $P(\overline{B} \mid \overline{A}) = 0.50 < 0.90$.

TABLE 2: A frequency distribution in which $A \rightsquigarrow B$ but $\overline{A} \not\rightsquigarrow \overline{B}$

Second attribute	First attribute A	$\overline{A}$	Total
B	95	10	105
$\overline{B}$	5	10	15
Total	100	20	120

Case 3PC. An empirical example, showing that $A \rightsquigarrow B \not\Rightarrow \overline{B} \rightsquigarrow \overline{A}$, is provided by Pauker and Pauker [**13**, pp. 291–292]. It was found that for a 30-year-old expectant mother, the probability of a Down-syndrome newborn is 1/885. Let A denote a normal fetus, and $\overline{A}$ an affected fetus, so that $P(A) = 884/885$ and $P(\overline{A}) = 1/885$. Amniocentesis is a highly accurate and reliable test, but it is not perfect: The results of 99.5% of the normal cases are indeed negative, and, similarly, 99.5% of the affected fetuses turn out positive results—in symbols, $P(B \mid A) = P(\overline{B} \mid \overline{A}) = 0.995$, where B denotes a negative test result (indicating normality). Combining these probabilities via Bayes' theorem yields $P(\overline{A} \mid \overline{B}) = 0.18$, to wit, a fetus may be diagnosed as Down by the test (i.e., $\overline{B}$ is observed), yet the posterior probability of its being affected is only 0.18. Hence, in this case $A \rightsquigarrow B$ does *not* imply $\overline{B} \rightsquigarrow \overline{A}$. Belief in the truth of 3PC was reported to result in major errors of probabilistic reasoning committed by physicians in drawing conclusions from clinical test results [**6**].

The F in 3PC goes contrary to many students' and researchers' intuitions, apparently because of departing from the Law of the Contrapositive, or the *modus tollens* syllogism that is affirmed by T in 3LI. *Modus tollens*, or contraposition is a powerful tool, often used in deductions, as in proof by contradiction. Extending this LI rule to PC inferences is seductive, and there is ample evidence showing that many think that what is true for certain deductions remains true for the analogous "almost certain" inferences. But as we saw, replacing the $\rightarrow$ of *modus tollens* by $\rightsquigarrow$ invalidates the inference. Belief in the validity of rule 3, applied to PC, is wrong! It is the *fallacy of probabilistic modus tollens.*

Moreover, the fallacy is too often responsible for misleading interpretations of statistical significance tests. In Harshbarger's textbook [**12**, p. 196], the logic of hypothesis testing is introduced as a form of denying the consequent (i.e., *modus tollens* or contraposition):

If A is true, then B must be true	(first premise)
B is not true	(second premise)
Therefore, A cannot be true	(conclusion)

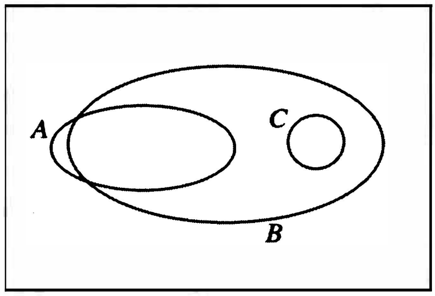
Up to now, this is correct—it is an instance of 3LI. The author explains further:

> Statistical hypothesis testing puts a slight twist on this logic, however, changing the deduction to read:
>
> If A is true, then B will probably be true
> B is not true
> ———
> Therefore, A is probably not true.

The conclusion in the bottom line is *false*. The analogy with the first paragraph breaks down when "must" is replaced by "probably," which refers in this case to a probability greater than 0.90 (i.e., to PC), since levels of significance are, as a rule, less than 0.10. In terms of hypothesis testing—letting A stand for "H_0" and B for "a test-statistic in the nonrejection region"—although $A \rightsquigarrow B$ holds, the assertion that this implies $\overline{B} \rightsquigarrow \overline{A}$ is not invariably true. This means that a test might yield a "significant" result and still not necessarily involve a high probability for the negation of H_0, as in the amniocentesis example, where A (the fetus is normal) stands for H_0 and $\overline{A}$ (the fetus is affected) for H_1. The erroneous interpretation of a significant result appears in many writings and is perpetuated widely in the teaching of statistics. I must confess to falling prey to this fallacy in the past [**7**]. Falk and Greenbaum [**10**] documented plenty of evidence showing that students and researchers wrongly interpret a significant result as a probabilistic confirmation of H_1.

Case 4LI. That $A \to B$ and $B \to C$ imply $A \to C$ is easily deduced by considering three sets that satisfy $A \subset B$ and $B \subset C$.

Case 4PS. An example showing that transitivity does not necessarily hold for probabilistic support is presented in FIGURE 2. Visual comparisons of the relative areas of the pertinent sets in the Venn diagram show that $P(B \mid A) > P(B)$ and $P(C \mid B) > P(C)$, whereas at the same time $P(C \mid A) = 0 < P(C)$.



$A \nearrow B$ and $B \nearrow C$, but $A \searrow C$

Figure 2 A counterexample to $A \nearrow B$ and $B \nearrow C \Rightarrow A \nearrow C$.

Case 4PC. That transitivity may also be violated for probabilistic confirmation is demonstrated in the following example [adapted from **14**, p. 79]: Let the universe include all animate and inanimate entities, and

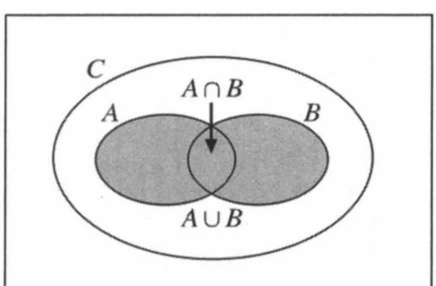
A = living humans

B = living organisms

C = nonmammals (including microorganisms).

These sets satisfy $A \rightsquigarrow B$ and $B \rightsquigarrow C$, whereas $P(C \mid A) = 0$, that is, $A \not\rightsquigarrow C$.

Cases 5LI and 6LI. The shaded area in FIGURE 3 represents $A \cup B$ and the overlapping area of A and B represents $A \cap B$. It is easily seen in the Venn diagram that $A \subset C$ and $B \subset C$ imply both $(A \cap B) \subset C$ (this holds also for disjoint events A and B since, by definition, ϕ is a subset of every set), thereby affirming the T in 5LI, and $(A \cup B) \subset C$, as maintained in 6LI.

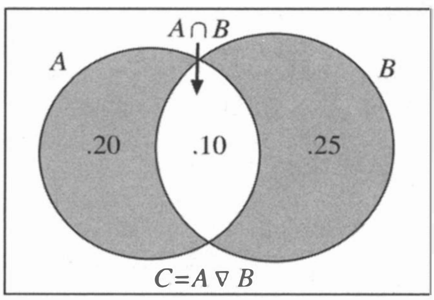


$A \to C$ and $B \to C \Rightarrow (A \cap B) \to C$ and $(A \cup B) \to C$

Figure 3 A involves C and B involves C imply $A \cap B$ involves C and $A \cup B$ involves C.

Case 5PS. The following story, adapted from Tribe [**16**, p. 1367] describes a legal situation in which the probability of a suspect's guilt, denoted C, has to be assessed: An armed robbery, taking 15 min to complete, was committed between 3:00 A.M. and 3:30 A.M. A witness saw the suspect in a car in the vicinity of the scene of the crime at 3:10 A.M. This evidence, denoted A, increases the suspect's initial probability of guilt (that had been based on rather shaky information). This is to say, $A \nearrow C$. Another witness saw the suspect in a car half a mile from the scene of the crime at 3:20 A.M. Denote this evidence B. By itself, $B \nearrow C$, but, taken together, the conjunction of the two testimonies, in effect, exonerates the suspect, to wit, $(A \cap B) \searrow C$.

This important point is also illustrated in FIGURE 4. The event C, represented by the shaded area in the diagram, is the *symmetric difference* of A and B, denoted $A \nabla B$ and defined by $A \nabla B = (A \cup B) - (A \cap B)$, which means either A or B but not both. Note that $P(C) = 0.45$, $P(C \mid A) = \frac{0.20}{0.30} > P(C)$, $P(C \mid B) = \frac{0.25}{0.35} > P(C)$, but $P(C \mid A \cap B) = 0 < P(C)$. Real-life examples of symmetric differences, that is, of exclusive disjunctions, are not easily found, but they do exist (e.g., [**8**, Problem 2.1.8]).



$A \nearrow C$ and $B \nearrow C$, but $(A \cap B) \searrow C$

Figure 4 A counterexample to $A \nearrow C$ and $B \nearrow C \Rightarrow (A \cap B) \nearrow C$.

Case 5PC. Let C be the symmetric difference of A and B. Suppose, now, that unlike the situation in FIGURE 4, the (nonempty) area of the intersection of A and B is less than 10% of that of either event. In this case, $P(C \mid A) > 0.90$, namely, $A \rightsquigarrow C$, and $P(C \mid B) > 0.90$, or $B \rightsquigarrow C$, but $P(C \mid A \cap B) = 0$, that is, $(A \cap B) \not\rightsquigarrow C$.

Case 6PS. TABLE 3 presents a counterexample (adapted from [**1**, p. 382]) to the claim that if A supports C and B supports C, then $A \cup B$ necessarily supports C. Ten items are simultaneously classified according to three attributes, each comprising two complementary categories. For a randomly drawn item, the following probabilities hold: $P(C) = 1/2$, $P(C \mid A) = 3/5 > P(C)$, $P(C \mid B) = 3/5 > P(C)$, and $P(C \mid A \cup B) = 3/7 < P(C)$. Thus, $A \nearrow C$ and $B \nearrow C$, whereas $(A \cup B) \searrow C$.

TABLE 3: $A \nearrow C$ and $B \nearrow C$, but $(A \cup B) \searrow C$

Second attribute	First attribute	
	A	$\overline{A}$
B	CCC	$\overline{C}\,\overline{C}$
$\overline{B}$	$\overline{C}\,\overline{C}$	$\overline{C}CC$

Case 6PC. The distribution of the simultaneous classification of 200 items into dichotomous categories of three attributes is presented in TABLE 4. The frequencies in the table refute the assertion that $A \rightsquigarrow C$ and $B \rightsquigarrow C$ invariably imply $(A \cup B) \rightsquigarrow C$, because $P(C \mid A) = 0.91 > 0.90$ and $P(C \mid B) = 0.91 > 0.90$, whereas $P(C \mid A \cup B) = 91/109 = 0.83 < 0.90$.

TABLE 4: $A \rightsquigarrow C$ and $B \rightsquigarrow C$, but $(A \cup B) \not\rightsquigarrow C$

Second attribute	First attribute		
	A	$\overline{A}$	Total
B	$91C$	$9\overline{C}$	100
$\overline{B}$	$9\overline{C}$	$9C$; $82\overline{C}$	100
Total	100	100	200

Conclusions

The analyses of the six rules of inference in TABLE 1 show that deductions concerning logical involvement and inferences concerning either probabilistic support or probabilistic confirmation often behave differently. Awareness of the similarities and, particularly, the dissimilarities between these modes of inference is essential both for teaching and for conducting research.

One important lesson from the cases of disaccord is that in research contexts, the consensually prevalent strategy of collecting converging evidence to strengthen the validity of a research hypothesis should be carefully checked in each case to make sure that it is not one of those odd cases, as in 5PS, in which $A \nearrow C$ and $B \nearrow C$, but $A \cap B \searrow C$. The very possibility that two items of evidence may each support a given event while their conjunction lowers the probability of the same event (as in the armed-robbery story)—even if rare—should serve to caution against routinely deciding that accumulation of evidence in favor of an uncertain event reinforces its validity.

In addition, whenever a statistically significant test result is obtained, although it is true that the research hypothesis (H_1) is supported, as in 3PS, it is *not* always true that H_1 becomes highly probable, as in 3PC. The "illusion of attaining improbability" [**10**, p. 78], that is, the faulty belief that rejection of the null hypothesis nearly refutes it, is very compelling and hard to eradicate. To uproot this tenacious misconception, teachers could include examples of the 3PC type in courses on statistical inference. This should alert students and investigators to the possibility that the credibility of H_1 has not reached a high enough level that justifies endorsing this hypothesis.

My impression is that the rules concerning logical involvements are psychologically acceptable to most of us more than their dissenting counterparts concerning probabilistic relations. The latter are often counterintuitive, if not bizarre. However, since most human experience takes place in an uncertain world, and certainty is hardly encountered, one could expect probabilistic rules to have a cognitive advantage. Chater and Oaksford [**3**] argue that probability theory rather than logic provides a more appropriate starting point for understanding human reasoning, and that logical deduction is a limiting case of probabilistic inference. Yet, erroneously believing in the truth of 3PC, 4PS, 4PC, 5PS, 5PC, 6PS, and 6PC, can reasonably be attributed to understanding the truth of the LI relations 3LI, 4LI, 5LI, and 6LI, and regarding the corresponding PS and PC relations as equivalent.

Tentatively, embracing the rules of logical inference and applying them to the probabilistic mode is a heuristic that yields good-enough answers in many cases, since the real situations in which the two systems differ are infrequent. As maintained by Tversky and Kahneman [**17**], reliance on heuristics is highly economical and usually effective. The heuristics are probabilistically reinforced in real situations, but they may lead to distinct errors on specific occasions. Though circumstances in which probabilistic inferences disobey the rules governing parallel logical deductions may be ecologically rare, these possibilities should not be dismissed. They signal a warning against unwittingly adhering to the good old logical prescripts when reasoning under uncertain conditions. Increased sensitivity to points of divergence between deductive and probabilistic inference should caution students, educators, and researchers against uncritically applying syllogistic rules to reasoning under uncertainty.

Readers are encouraged to find more, perhaps better, demonstrations, proofs, and counterexamples for the 18 combinations. For a thorough discussion of the vagaries of relations between events see Carnap [**1**, chap. 6]. More analogies and misanalogies between principles of logical and probabilistic inference can be found in Falk and Bar-Hillel [**9**], and Salmon [**15**].

Acknowledgments. The author thanks Raphael Falk for thoughtful advice and technical help, and the editor and an anonymous reviewer for constructive suggestions of how to crystallize the exposition. Yael Oren provided the final touch to the figures.

REFERENCES

1. R. Carnap, *Logical Foundations of Probability*, University of Chicago Press, Chicago, 1950.
2. R. P. Carver, The case against statistical significance testing, *Harvard Educational Review* **48**(3) (1978) 378–399.
3. N. Chater and M. Oaksford, The probability heuristics model of syllogistic reasoning, *Cognitive Psychology* **38** (1999) 191–258.
4. K. L. Chung, On mutually favorable events, *Annals of Mathematical Statistics* **13** (1942) 338–349.
5. I. M. Copi, *Introduction to Logic*, 3rd ed., Macmillan, London, 1968.
6. D. M. Eddy, Probabilistic reasoning in clinical medicine: Problems and opportunities, in *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky (eds.), Cambridge University Press, Cambridge, 1982, pp. 249–267.
7. R. Falk, On coincidences, *The Skeptical Inquirer* **6**(2) (1981–82) 18–31.
8. R. Falk, *Understanding Probability and Statistics: A Book of Problems*, A K Peters, Wellesley, MA, 1993.
9. R. Falk and M. Bar-Hillel, Probabilistic dependence between events, *Two-Year College Mathematics Journal* **14**(3) (1983) 240–247.
10. R. Falk and C. W. Greenbaum, Significance tests die hard: The amazing persistence of a probabilistic misconception, *Theory & Psychology* **5**(1) (1995) 75–98.
11. S. Gudder, Do good hands attract? this MAGAZINE **54**(1) (1981) 13–16.
12. T. R. Harshbarger, *Introductory Statistics: A Decision Map*, Macmillan, New York, 1977.
13. S. P. Pauker and S. G. Pauker, The amniocentesis decision: An explicit guide for parents, in *Birth Defects: Original Article Series: Risk, Communication, and Decision Making in Genetic Counseling*, C. J. Epstein, et al. (eds.), The National Foundation, New York, 1979, Vol. 15, pp. 289–324.
14. W. C. Salmon, Confirmation, *Scientific American* **228**(5) (1973) 75–83.
15. W. C. Salmon, Confirmation and relevance, in *Induction, Probability, and Confirmation*, G. Maxwell and R. M. Anderson Jr. (eds.), University of Minnesota Press, Minneapolis, 1975, pp. 3–36.
16. L. H. Tribe, Trial by mathematics: Precision and ritual in the legal process, *Harvard Law Review* **84**(6) (1971) 1329–1393.
17. A. Tversky and D. Kahneman, Judgment under uncertainty: Heuristics and biases, *Science* **185** (1974) 1124–1131.

Wazir Circuits on an Obstructed Chessboard

IRL C. BIVENS
Davidson College
Davidson, NC 28035
irbivens@davidson.edu

ARTHUR L. HOLSHOUSER
Charlotte, NC 28208

BENJAMIN G. KLEIN
Davidson College
Davidson, NC 28035
beklein@davidson.edu

A "wazir" is a now obsolete chess piece that moves exactly one square at a time vertically or horizontally. It is sometimes referred to as a "single-step rook". Although its role in the game of chess has long since been supplanted by the more versatile rook, the wazir still makes an occasional appearance in the chess variants known as "fairy chess". In addition, the wazir provides us with a useful vehicle for the discussion of some interesting results about certain rook paths. For example, consider the following problem that appeared on the 1999 International "Tournament of the Towns" competition:

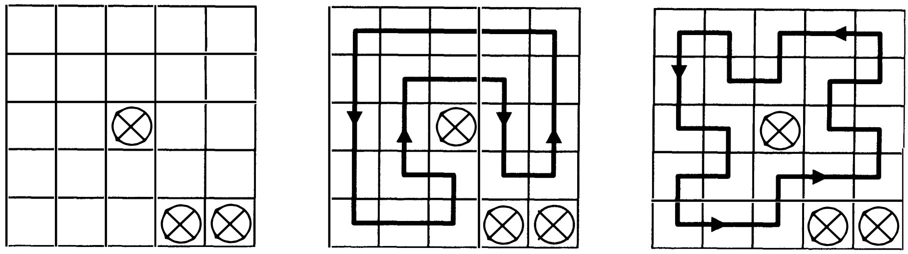
> A move of the rook consists of passing to a neighboring cell in either the horizontal or the vertical direction. After 64 moves the rook visited all cells of the 8×8 chessboard and returned back to the initial cell. Prove that the number of moves in the vertical direction and the number of moves in the horizontal direction are distinct.

Since the rook is required to move a single square at a time, this problem has a natural formulation in terms of the wazir. (We will solve this problem shortly. The reader may wish to think at least briefly about some of the issues involved before proceeding.)

In this article, and in the accompanying exercises, we will discuss a parity result for wazir circuits on chessboards of arbitrary size. We allow our chessboards to be "obstructed" in the sense that movement into certain squares is forbidden. (Imagine squares that are occupied by chess pieces the same color as the wazir.) For example, FIGURE 1(a) shows a 5×5 chessboard with three forbidden squares. A square into which movement is forbidden will be called "obstructed." Any unobstructed square will be called "open." A "wazir circuit" is a sequence of moves in which the wazir begins in an open square, ends in the same square, and visits every other open square on the board exactly once. Our goal is to determine the parity of the number of moves the wazir makes in a specified direction (e.g., to the right) during a circuit. The answer will be the result of an interesting interplay between parity considerations and the "geometry" of the circuit.

Connecting parity with geometry

The existence of a wazir circuit on a chessboard implies that any two open squares of the board can be connected by a path of the wazir. For this reason, we will not consider obstructed boards in which the open squares are separated into disconnected



(a) An obstructed board with three forbidden squares

(b) A circuit with five moves to the right and six upward moves

(c) A circuit with six moves to the right and five upward moves

Figure 1 An obstructed chess board along with two wazir circuits.

components by the obstructed squares. Furthermore, in order for a wazir to begin and end a circuit in the same square, the number of moves to the right must equal the number of moves to the left and the number of upward moves must equal the number of downward moves. It follows that the total number of moves in a circuit must be even. Since the number of moves in a circuit is equal to the number of open squares on the board, we will restrict our attention to boards with an even number of open squares. On such a board, the parity of the total number of open squares that belong to alternating rows of the board is well-defined, independently of the choice of alternate rows. If this number is odd (even), we will say that the board has an odd (even) "horizontal splitting." Similarly, by considering columns instead of rows, we can define an odd (even) vertical splitting. For example, the board in FIGURE 1(a) has an even horizontal splitting and an odd vertical splitting. The existence of a wazir circuit on a board constrains the types of splittings that can occur. For example, if a board has a wazir circuit, the parity of a horizontal splitting of the board is the same as that for a vertical splitting of the board, if and only if the total number of open squares on the board is divisible by 4. (See part (c) of Exercise 1. The existence of a wazir circuit also has other important implications. See Exercise 2 for an application of wazir circuits to the problem of "tiling" a board with dominoes.)

In general, the parity of the number of moves by the wazir in a specified direction during a circuit is not determined by the board configuration alone. For example, FIGURE 1(b) depicts a wazir circuit with 5 moves to the right and 6 upward moves, while FIGURE 1(c) shows a circuit of the same board with 6 moves to the right and 5 upward moves. Our aim is to understand these changes in parity. As illustrated in FIGURE 1, we associate to any wazir circuit an oriented "circuit polygon", which is a directed graph whose vertices are the centers of open squares and whose edges indicate the moves of the wazir. We will say that any obstructed square that is within the region bounded by the circuit polygon is "inside" the polygon, while any other obstructed square is "outside" the circuit polygon. Note that in the circuit of FIGURE 1(b), all the obstructed squares lie *outside* the circuit polygon, while in the circuit of FIGURE 1(c), two obstructed squares are outside the polygon and one is *inside*. This observation is important. We will see that the board configuration, together with the number of obstructed squares inside the circuit polygon, determine the parity of the number of moves in a given direction. For example, in any circuit of the board in FIGURE 1, the number of moves to the right by the wazir will have parity opposite that of the number of obstructed squares inside the circuit polygon.

THEOREM. *Suppose that a chessboard possesses a wazir circuit. If the board has an odd horizontal splitting then the number of moves to the right (left) in a wazir* *circuit has the same parity as the number of obstructed squares inside the circuit polygon. If a board has an even horizontal splitting, the number of moves to the right (left) has parity opposite to that of the number of obstructed squares inside the circuit polygon. [Replacing "horizontal splitting" by "vertical splitting" and "moves to the right (left)" by "upward (downward) moves" gives the corresponding parity result for vertical moves.]*

(As might be expected, there is a close relationship between wazir circuits and rook circuits. Exercise 7 provides a reformulation of this theorem in terms of rook circuits.)

The concept of the "parity" of a horizontal splitting allows us to refer to the effect of the board configuration in a natural way. In addition, this concept makes it convenient to apply our theorem repeatedly to different wazir circuits on the same board. However, for the purpose of proving the theorem, it will be useful to have a formulation that doesn't made a distinction between two types of splittings. To do this, let us suppose that a chessboard possesses a wazir circuit. Number the rows of the board from bottom to top, and let E denote the number of open squares on even-numbered rows. Given a wazir circuit, let R denote the number of moves to the right by the wazir, and let I denote the number of obstructed squares inside the circuit polygon. In terms of these parameters, it is clear that our theorem is equivalent to the equation

$$R + I + E \equiv 1 \pmod 2 \tag{1}$$

Consequently, to prove the theorem, it suffices to prove equation (1). (For the conclusions about vertical moves, simply rotate the board 90 degrees. Although equation (1) appears to have little to do with area considerations, its proof includes a surprising appearance by Pick's Theorem.) Before proceeding to the proof, we will consider some simple applications of the theorem.

Some applications

Examples are given in [**1**] of chessboards that possess a unique wazir circuit (up to reversing the direction of all the wazir moves). On such a board, the actual *number* of moves in any specified direction during the circuit is uniquely determined. We can use our theorem to prove this is also the case for the board in FIGURE 2, even though the circuit depicted is *not* unique. (See Exercise 4.)

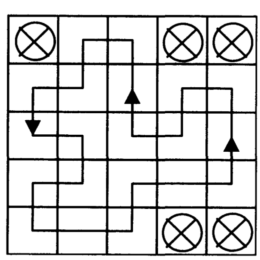


Figure 2 Every wazir circuit has five moves in each direction.

Note that in the circuit shown, there are 5 moves by the wazir in each of the 4 possible directions. To see why this follows from our theorem, let R denote the number of moves to the right and let U denote the number of moves upward in a wazir circuit of this board. In a circuit the total number of moves $2R + 2U$ must equal the number of open squares on the board. We conclude that

$$2R + 2U = 20 \qquad \text{or} \qquad R + U = 10$$

Furthermore, we must have $R, U \geq 4$ in order for the wazir to get from the left side of the board to the right side, or from the bottom of the board to the top. Since the board has an even horizontal splitting, and no obstructed square can lie inside a circuit polygon, we conclude from the theorem that R (and therefore U) must be odd. But then $R = U = 5$ which justifies our claim.

Note that for the circuit in FIGURE 2, the wazir moves the same number of times horizontally as it does vertically. In general, the number of horizontal moves during a wazir circuit need not be the same as the number of vertical moves, and for many boards *must* be different. For example, let us return to the problem on the 1999 International "Tournament of the Towns" competition that we introduced at the beginning of this article. In terms of the wazir, this problem becomes:

Prove that for any wazir circuit on an unobstructed 8×8 chessboard, the number of moves in the vertical direction and the number of moves in the horizontal direction must be distinct.

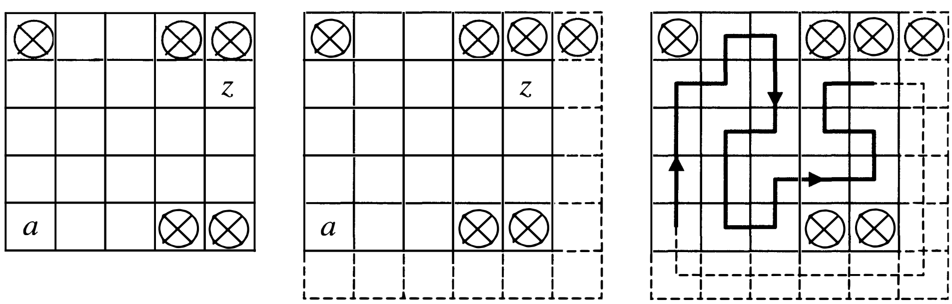
Let R denote the number of moves to the right by the wazir. The 8×8 chessboard has an even horizontal splitting and no obstructed squares. Therefore, it follows from our theorem that R is odd. The total number of horizontal moves of the wazir is $2R$. Were the number of moves in the vertical direction equal to the number of moves in the horizontal direction, we would have $2R = 32$, which would imply that R is 16. We conclude that the number of moves in the vertical direction and the number of moves in the horizontal direction must be distinct.

More generally, a wazir circuit on any unobstructed $2k \times 2k$ chessboard must have twice an *odd* number of horizontal (vertical) moves. (Equivalently, the number of horizontal moves, and the number of vertical moves, must be congruent to 2 modulo 4.) Consequently, no circuit with equal numbers of horizontal and vertical moves exists if k is even. (If k is odd then the theorem does not preclude the existence of a circuit with equal numbers of moves in the horizontal and vertical directions. In fact, it is not hard to show that in this case such a circuit is always possible. See Exercise 5.)

(a) A tour with two moves to the right (b) A tour with three moves to the right

Figure 3 Two tours with different parities for the number of moves to the right.

In a wazir "tour" of a board, the wazir visits each square exactly once but is not required to begin and end its trip in the same square. Suppose we specify the beginning and ending squares of the tour. Can the parity of the number of moves in a given direction then be determined from the configuration of the board? In general the answer is "no", even for very simple boards. For example, there are two moves to the right for the tour in FIGURE 3(a) and three moves to the right for the tour in FIGURE 3(b). On the other hand, we can apply our theorem to prove that for every tour from square a to square z of the board in FIGURE 4(a), the number of moves upward will be odd, while the number of moves in any other direction will be even. To prove this, we adjoin a column to the right and a row at the bottom of the board as in FIGURE 4(b). Any tour from a to z in the original board can be completed to a wazir circuit with the addition of edges in the adjoined column and row. FIGURE 4(c) illustrates this completion in the case of one particular tour. The modified board has an even horizontal splitting, and for any tour from a to z that is completed to a circuit, the pair of obstructed squares in the lower right-hand corner of the board will lie inside the corresponding circuit polygon. Therefore, the number of moves to the right (left) in the circuit must be odd. Since completing the tour to a circuit adds one move right and 5 moves left, we conclude that the original tour must have had an even number of moves to the right and to the left. Similarly, the augmented board has an odd vertical splitting, from which it follows that the number of moves upward (downward) in the completed circuit will be even. Since completing the tour to a circuit adds one move upward and 4 moves downward, the original tour must have had an odd number of upward moves and an even number of downward moves. (A careful argument shows that a tour of this board has either 8 moves to the right, 4 moves to the left, 5 moves upward, and 2 moves downward, or 6 moves to the right, 2 moves to the left, 7 moves upward, and 4 moves downward. See part (a) of Exercise 6.)



(a) Tours are to start in square a and end in square z

(b) Adjoining a row and a column

(c) Completing a tour to a circuit

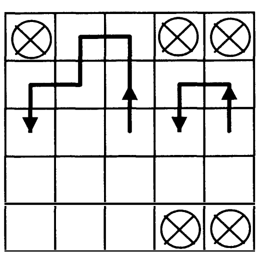
Figure 4 For every tour from a to z, the number of upward moves is odd and the number of moves in any other direction is even.

The proof

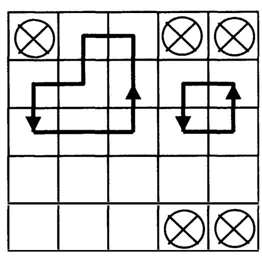
We now proceed to the proof of our theorem. A "chessboard" will be a grid that consists of $m \geq 2$ rows and $n \geq 2$ columns, with the rows numbered from bottom to top. We let L_i, $i = 1, 2, \ldots, m-1$ denote the top edge of row i of the board. The number of open squares that lie above L_i will be denoted by S_i. "Corner points" will refer to the corners of squares of the board, both open and obstructed. Suppose a board has a wazir circuit. We let U_i denote the number of upward edges in the circuit polygon that intersect L_i and we let C_i denote the number of corner points on L_i that belong to the interior of the circuit polygon. (For example, in FIGURE 2 with $i = 3$, we have $S_3 = 7$, $U_3 = 2$, and $C_3 = 3$.) A result that is key to our argument is the fact that

$$C_i + S_i + U_i \equiv 0 \pmod 2, \quad i = 1, 2, \ldots, m-1 \tag{2}$$

To prove (2), let us suppose first that $U_i = 0$. If $U_i = 0$ then all open squares are on one side of L_i. Therefore, $C_i = 0$ and either $S_i = 0$ or S_i is equal to the total number of open squares. Since we are assuming the existence of a wazir circuit on the board, the total number of open squares is even. Thus, S_i is even and (2) is valid. Suppose next that U_i is positive and assume without loss of generality that the circuit polygon is oriented counterclockwise. The interior of the circuit polygon will then be (locally) to the right of any downward edge, to the left of any upward edge, above any edge directed to the right, and below any edge directed to the left. (It follows that as we move from left to right along L_i, the down and up edges we encounter alternate, beginning with a down edge and ending with an up edge.) First, we "truncate" the wazir circuit at row i by removing from the circuit all horizontal edges between squares on rows i or below, and by removing all vertical edges that either intersect L_{i-1} or lie below L_{i-1}. We also make all squares in row i open. What remains will be that portion of the circuit polygon in rows above row i, together with the vertical edges between row i and row $i+1$. Part (a) of FIGURE 5 illustrates this truncation procedure in row 3 for the circuit in FIGURE 2. Next, for each downward edge terminating in row i, we adjoin enough horizontal edges directed to the right to reach the first vertical edge to the right of the downward edge. Note that one horizontal edge will be added for each corner point on L_i that belongs to the interior of the circuit polygon. As a consequence, the resulting total number of horizontal edges in row i will be C_i. What results will be one or more (disjoint) simple oriented closed polygons. (Strictly speaking, these "polygons" should be interpreted as directed graphs.) For example, in FIGURE 5(b) two such polygons are produced. Since the number of edges for each closed polygon must be even, it follows that the total number of edges, over all the polygons, is even. These edges consist of the C_i horizontal edges in row i, the U_i upward vertical edges that exit a square in row i and the remaining S_i edges that exit an open square above row i. But then, $C_i + S_i + U_i$ is even, which proves (2).



(a) Truncating the circuit of FIGURE 2 in row three



(b) Producing two polygons from the truncated circuit

Figure 5 Truncating a wazir circuit and constructing closed polygons.

Summing identity (2) over $i = 1, \ldots, m-1$ yields

$$0 \equiv \sum_{i=1}^{m-1}(C_i + S_i + U_i) = \sum_{i=1}^{m-1} C_i + \sum_{i=1}^{m-1} S_i + \sum_{i=1}^{m-1} U_i \pmod 2 \qquad (3)$$

Note that $C = \sum_{i=1}^{m-1} C_i$ is the total number of corner points in the interior of the circuit polygon and that $U = \sum_{i=1}^{m-1} U_i$ is the total number of upward directed edges in the polygon. Now in the sum, $\sum_{i=1}^{m-1} S_i$, open squares on even-numbered rows are counted an odd number of times while open squares on odd-numbered rows are counted an even number of times. It follows that $\sum_{i=1}^{m-1} S_i$ has the same parity as the total number of open squares that belong to the even-numbered rows of the board. If we let E denote the number of open squares on even-numbered rows, then

$$E \equiv \sum_{i=1}^{m-1} S_i \pmod 2$$

and equation (3) becomes

$$0 \equiv C + E + U \pmod 2 \tag{4}$$

Note that equation (4) is already a parity invariant for wazir circuits. However, the parameter C is of less intrinsic interest than the number of obstructed squares inside the circuit polygon. To complete the proof of our theorem, we count the number C in a different way. Assume without loss of generality that the squares of the board are unit-squares. We can then "embed" our board in the xy-plane so that the sides of the board are parallel to the coordinate axes and so that the centers of squares of the board fall on lattice points of the plane. Given a circuit polygon, the vertical and horizontal lines through lattice points will partition the interior of the polygon into unit squares, each of which is centered at a corner point. It follows that the area of the polygon will be numerically equal to the number C. As we did in equation (1), let R denote the number of right-directed horizontal edges of the circuit polygon, and let I denote the number of obstructed squares inside the polygon. Recall from Pick's theorem **[2]** that the area enclosed by a simple polygon in the plane whose vertices are lattice points is given by the formula

$$\frac{1}{2}p + q - 1$$

where p is the number of lattice points on the polygon and q is the number of lattice points within the interior of the polygon. For us, $p = 2U + 2R$ and $q = I$, so that

$$C = \frac{1}{2}(2U + 2R) + I - 1 = U + R + I - 1$$

Substituting this value of C into equation (4) yields

$$\begin{aligned} 0 &\equiv C + E + U = (U + R + I - 1) + E + U \\ &= 2U + R + I + E - 1 \equiv R + I + E - 1 \pmod 2 \end{aligned}$$

and thus

$$R + I + E \equiv 1 \pmod 2$$

But this is equation (1), which completes the proof of our theorem.

Exercises

We conclude with some exercises for the reader.

1. Suppose that a chessboard is 2-colored in the usual way and that we number the rows from bottom to top, and the columns from left to right. Let O_r denote the total number of open squares that belong to odd-numbered rows, and let O_c denote the total number of open squares that belong to odd-numbered columns.

(a) Show that for one color, each open square of that color is counted exactly once in the sum $O_r + O_c$, while each open square of the opposite color is counted either not at all, or twice, in this sum.

(b) Assume the board has a wazir circuit. Explain why the number of open squares of one color is equal to the number of open squares of the opposite color.

(c) Use (a) and (b) to prove that if the board has a wazir circuit, the parity of a horizontal splitting of the board is equal to that of a vertical splitting of the board, if and only if the total number of open squares on the board is divisible by 4.

2. Assume that a (possibly obstructed) chessboard possesses a wazir circuit.
   (a) Prove that it is possible to "tile" the open squares of the board with 2×1 dominoes.
   (b) Pick any two open squares of different colors on the board and make them both obstructed. Prove that it is possible to tile the remaining open squares of the board with dominoes.

   (These results are essentially a restatement of Gomory's Theorem [**3**].)
3. Consider an 8×8 chessboard with exactly 2 obstructed squares that are located in diagonally opposite corners of the board. A well-known puzzle is to prove that such a board has no wazir circuit.
   (a) Use the result of Exercise 1(b) to solve this puzzle. (This is the "usual" solution.)
   (b) Use the result of Exercise 1(c) to give a different solution to this puzzle.
4. Prove that the board in Figure 2 has three wazir circuits (up to reversing the direction of all the wazir moves).
5. Prove that if k is odd, every $2k \times 2k$ unobstructed chessboard has a wazir circuit with $2k^2$ horizontal moves and $2k^2$ vertical moves. (*Hint:* Justify the steps in the following sketch of an inductive proof. After handling the base case, suppose the result is true for a $2k \times 2k$ board B with a wazir circuit that begins in the upper right-hand square a of B with a move to the *left* into square b. Embed B in the center of a $2(k+2) \times 2(k+2)$ board and move the wazir from a to b via the following sequence: 1 move up, 2 moves to the right, 1 move up, $2k+3$ moves to the left, $2k+3$ downward moves, $2k+3$ moves to the right, $2k+1$ upward moves, 1 move to the left, $2k$ downward moves, $2k+1$ moves to the left, $2k+1$ upward moves, $2k-1$ moves to the right, 1 move down to square b. Then, follow the wazir circuit of B from square b back to square a.)
6. (a) Prove that every wazir tour from square a to square z of the board in Figure 4(a) has either 8 moves to the right, 4 moves to the left, 5 moves upward, and 2 moves downward, or 6 moves to the right, 2 moves to the left, 7 moves upward, and 4 moves downward.
   (b) Show that there are six wazir tours from square a to square z of the board in Figure 4(a), two of which have 8 moves to the right.
7. A rook circuit is a sequence of moves in which the rook begins in an open square, ends in the same square, and enters every other open square on the chessboard exactly once. As the rook travels around the board it enters an open square in one of (at most) four directions. Prove that our theorem may be restated in terms of rook circuits as follows:

   Suppose that a chessboard possesses a rook circuit. If the board has an odd horizontal splitting, the number of squares entered from the left (right) during *the circuit has the same parity as the number of obstructed squares inside the circuit polygon. If the board has an even horizontal splitting, the number of squares entered from the left (right) during the circuit has parity opposite to that of the number of obstructed squares inside the circuit polygon. [Replacing "horizontal splitting" by "vertical splitting" and "from the left (right)" by "from below (above)" gives the corresponding parity result for vertical rook moves.]*

REFERENCES

1. Richard Guy and Mark Paulhus, Unique rook circuits, this MAGAZINE **75** (2002) 380–387.
2. J. Smart, *Modern Geometries*, 5th ed., Brooks/Cole, 1998.
3. John Watkins, *Across the Board: The Mathematics of Chessboard Problems*, Princeton University Press, 2004

NOTES

What if Archimedes Had Met Taylor?

JASON SLOWBE
San Marcos High School
1615 San Marcos Blvd.
San Marcos, CA 92078
jslowbe@euhsd.k12.ca.us

Even before Archimedes, mathematicians knew that a circle's diameter is proportional to its circumference, and that the area of a circle is proportional to the square of its radius. It was Archimedes, however, who first supplied a rigorous proof that these two proportionality constants were the same; we now call them π (see [**1**, p. 31]). He also showed that increasing the number of sides of regular polygons inscribed and circumscribed about a circle creates perimeters that become arbitrarily close to the circle's perimeter. Starting with inscribed and circumscribed regular hexagons about a circle of unit diameter, then doubling the number of sides to 12, 24, 48, and 96, Archimedes calculated their perimeters and thus created lower and upper endpoints of an interval that shrinks in length to zero yet always contains the circle's perimeter known to be π. This method, however, does not produce a single *numeric* approximation of π. Is there some way to improve his method's ability to generate actual digits of π? The answer is *yes*, and Taylor will lend a hand.

It is thought that Archimedes also considered *areas* of these inscribed and circumscribed polygons in his quest to prove that the constants derived from the perimeters and areas of circles were the same [**2**, p. 281]. Increasing n "fills the unit circle," thereby producing a similar algorithm for generating decimal approximations of π. We will begin by analyzing the areas-based algorithm. This leads to a more efficient way to approximate π's value. An analogous perimeters-based result further improves its efficiency. We will see several nice connections between the two methods.

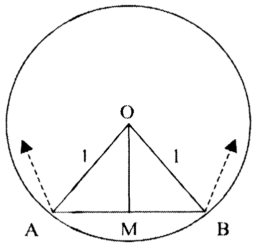
Pi filling, Archimedes style!

Archimedes created an algorithm that doubled the number of sides of inscribed and circumscribed polygons to create lower and upper bounds that both converge onto the actual value of π as $n \to \infty$. We start by finding the area of an arbitrary regular n-gon inscribed in the unit circle using Area $= (1/2)$perimeter $\cdot$ altitude, where the altitude is the segment $\overline{OM}$ extending from the center of the circle and perpendicular to any side of the regular n-gon (FIGURE 1).

We have established the lower bound, so it remains to define an expression for the upper bound from the circumscribed regular n-gon (FIGURE 2).

Evaluating the intervals for $n = 6, 12, 24, 48, 96$ is straightforward. As n increases, the bounds from the intervals converge to π.

Archimedes' method generates intervals that capture π. *Can his method be improved to generate a decimal approximation to π?* Knowing π to be ≈ 3.14159265, inspection shows that the circumscribed areas are closer to π than are the inscribed



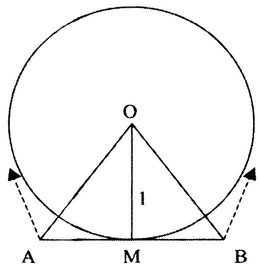
$$m\angle AOB = \frac{2\pi}{n} \qquad m\angle AOM = \frac{1}{2}\cdot\frac{2\pi}{n} = \frac{\pi}{n}$$

$$OM = \cos\left(\frac{\pi}{n}\right) \qquad AB = 2\sin\left(\frac{\pi}{n}\right)$$

$$\text{Area } n\text{-gon} = \frac{1}{2}\text{perimeter}\cdot\text{altitude}$$

$$= \frac{1}{2}\left(n\cdot 2\sin\left(\frac{\pi}{n}\right)\right)\left(\cos\left(\frac{\pi}{n}\right)\right) = n\sin\left(\frac{\pi}{n}\right)\cos\left(\frac{\pi}{n}\right)$$

Figure 1



$$m\angle AOB = \frac{2\pi}{n} \qquad m\angle AOM = \frac{1}{2}\cdot\frac{2\pi}{n} = \frac{\pi}{n} \qquad AB = 2\tan\left(\frac{\pi}{n}\right)$$

$$\text{Area } n\text{-gon} = \frac{1}{2}\text{perimeter}\cdot\text{altitude}$$

$$= \frac{1}{2}\left(n\cdot 2\tan\left(\frac{\pi}{n}\right)\right)(1) = n\tan\left(\frac{\pi}{n}\right)$$

Figure 2

areas, but how much closer? It is straightforward to calculate "π's relative location in each interval": Divide the distance from the lower bound to π by the total length of the interval (TABLE 1).

TABLE 1: Location of π in corresponding intervals

n-gon	Interval	Ratio into interval $\dfrac{\pi - \text{lower bound}}{\text{upper bound} - \text{lower bound}}$
6	(2.598076, 3.464102)	.627599
12	(3.000000, 3.215390)	.657377
24	(3.105829, 3.159660)	.664373
48	(3.132629, 3.146086)	.666095
96	(3.139350, 3.142715)	.666524

There seems to be a trend in the ratios, perhaps approaching $2/3$ as n increases, but a limit may or may not exist. Another problem: this numeric result requires prior knowledge of the value of π. Fortunately, these problems turn out to be relatively minor.

Archimedes, meet my friend Taylor

Define $f(n)$ for the ratio associated with π's location in each interval, as in TABLE 1.

$$f(n) = \frac{\pi - n\sin\left(\frac{\pi}{n}\right)\cos\left(\frac{\pi}{n}\right)}{n\tan\left(\frac{\pi}{n}\right) - n\sin\left(\frac{\pi}{n}\right)\cos\left(\frac{\pi}{n}\right)} = \frac{\cos\left(\frac{\pi}{n}\right)\left[2\pi - n\sin\left(\frac{2\pi}{n}\right)\right]}{2n\sin^3\left(\frac{\pi}{n}\right)}.$$

Using Taylor series and some calculus-based reasoning, we can translate our hunch into a proof that the limit is in fact 2/3. L'Hôpital's rule lets us consider only the leading terms, as all of the other terms have higher degrees and do not influence the limit.

$$\lim_{n\to\infty} \frac{\cos\left(\frac{\pi}{n}\right)\left[2\pi - n\sin\left(\frac{2\pi}{n}\right)\right]}{2n\sin^3\left(\frac{\pi}{n}\right)} = \lim_{n\to\infty} \cos\left(\frac{\pi}{n}\right) \cdot \lim_{n\to\infty} \frac{\left(2\pi - n\left[\frac{2\pi}{1!n} - \frac{(2\pi)^3}{3!n^3} + \cdots\right]\right)}{2n\left(\frac{\pi}{1!n} - \frac{\pi^3}{3!n^3} + \cdots\right)^3}$$

$$= \lim_{n\to\infty} \frac{\frac{4\pi^3}{3n^2} + O\left(\frac{1}{n^4}\right)}{\frac{2\pi^3}{n^2} + O\left(\frac{1}{n^4}\right)} = \frac{2}{3}.$$

The fascinating part is that no knowledge of the numeric value of π is required! The numeric value of π is located approximately 2/3 the way into any given interval, and becomes arbitrarily close to being exactly 2/3 as n goes to infinity. If Taylor had met Archimedes, they could have collaborated to develop a remarkably more efficient method of approximating the digits of π. *How much better?*

TABLE 2: Estimating π from Areas Intervals

n-gon	Estimates for π $\frac{1}{3} \cdot \left[\frac{n}{2}\sin\left(\frac{2\pi}{n}\right)\right] + \frac{2}{3} \cdot \left[n\tan\left(\frac{\pi}{n}\right)\right]$
6	**3.**175 426 48
12	**3.143** 593 54
24	**3.141** 716 14
48	**3.141 6**00 34
96	**3.141 593** 13

Adding 2/3 of the length of each interval to the lower bound gives improved approximations for π. This is equivalent to adding 1/3 of the lower bound and 2/3 of the upper bound (TABLE 2). This result improves Archimedes' method, which for $n = 96$ merely bounded π between $3\frac{10}{71} < \pi < 3\frac{1}{7}$, or 3.14084507 and 3.14285714. However, Archimedes used perimeters, not areas, in calculating these fractional bounds. Is there a similar trend for π's location in the intervals formed from the perimeters?

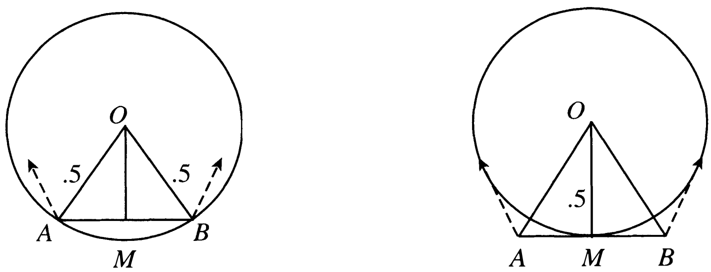
Perimeters

We can ask a related question: How do the perimeters-based intervals capture π? Making a table like TABLE 1 reveals another apparent trend to the ratios—this time, approaching 1/3. Will our luck hold?

For the areas-based approach, we had to use a unit circle with area π. For a circle's perimeter to be π, it must have unit diameter, so the inscribed and circumscribed regular polygons must be constructed about a circle with diameter 1. Increasing n generates polygons whose perimeters approach the circle's perimeter, which we know is π. The inscribed and circumscribed perimeters form the lower and upper bounds for π (FIGURE 3).

Perimeter of Inscribed n-gon

Perimeter of Circumscribed n-gon



$$m\angle AOM = \frac{\pi}{n}$$

$$\text{Perimeter } n\text{-gon} = n \cdot 2 \cdot \left(\frac{1}{2} \cdot \sin\left(\frac{\pi}{n}\right)\right) = n \sin\left(\tfrac{\pi}{n}\right)$$

$$m\angle AOM = \frac{\pi}{n}$$

$$\text{Perimeter } n\text{-gon} = n \cdot 2 \cdot \left(\frac{1}{2} \cdot \tan\left(\frac{\pi}{n}\right)\right) = n \tan\left(\frac{\pi}{n}\right)$$

Figure 3

Again, define $f(n)$ to give the ratio associated with π's location in each interval, and use Taylor series to get another remarkable result:

$$\lim_{n\to\infty} \frac{\pi - n\sin\left(\frac{\pi}{n}\right)}{n\tan\left(\frac{\pi}{n}\right) - n\sin\left(\frac{\pi}{n}\right)} = \lim_{n\to\infty} \frac{\cos\left(\frac{\pi}{n}\right)\left[\pi - n\sin\left(\frac{\pi}{n}\right)\right]}{n\sin\left(\frac{\pi}{n}\right) - \frac{n}{2}\sin\left(\frac{2\pi}{n}\right)}$$

$$= \lim_{n\to\infty} \frac{\pi - n\left(\frac{\pi}{1!n} - \frac{\pi^3}{3!n^3} + \cdots\right)}{n\left(\frac{\pi}{1!n} - \frac{\pi^3}{3!n^3} + \cdots\right) - \frac{n}{2}\left(\frac{2\pi}{1!n} - \frac{(2\pi)^3}{3!n^3} + \cdots\right)}$$

$$= \lim_{n\to\infty} \frac{\frac{\pi^3}{3!n^2} + O\left(\frac{1}{n^4}\right)}{-\frac{\pi^3}{3!n^2} + \frac{(2\pi)^3}{2\cdot 3!n^2} + O\left(\frac{1}{n^4}\right)} = \lim_{n\to\infty} \frac{\frac{\pi^3}{6n^2}}{\frac{6\pi^3}{12n^2}} = \frac{1}{3}$$

Consider the new decimal approximations for π being $1/3$ the way into the intervals of perimeters (TABLE 3).

TABLE 3: Improving the Estimates for π

n-gon	Interval	Estimates for π: $\frac{2}{3} \cdot n\sin\left(\frac{\pi}{n}\right) + \frac{1}{3} \cdot n\tan\left(\frac{\pi}{n}\right)$
6	(3.000 000, 3.464 102)	**3.1**54 700 54
12	(3.105 829, 3.215 390)	**3.142** 349 13
24	(3.132 629, 3.159 660)	**3.141 6**39 05
48	(3.139 350, 3.146 086)	**3.141 595** 54
96	(3.141 032, 3.142 715)	**3.141 592** 83

Thanks to Taylor, Archimedes' method gives an improved approximation, accurate to six places, by looking $1/3$ the way into the interval for the 96-gon. Applying the $1/3$ result to Archimedes' fractional interval $3\frac{10}{71} < \pi < 3\frac{1}{7}$ gives $\left(\frac{2}{3}\right)\left(3\frac{10}{71}\right) + \left(\frac{1}{3}\right)\left(3\frac{1}{7}\right) = \frac{4684}{1491} \approx 3.14151576$. While less accurate than the $n = 24, 48$, and 96 approximations in

TABLE 3, looking $1/3$ the way into Archimedes' interval gives a tremendous improvement for approximating π compared to merely giving an interval known to contain π. (Never mind that it is accurate to four places!) Such a simple and elegant improvement gives much greater precision to one of the greatest tasks of antiquity—pinning down the mysterious value of π.

Connecting the area and perimeter methods

By considering the perimeters of inscribed and circumscribed regular polygons in a circle with perimeter π, Archimedes obtained the fractional interval $3\frac{10}{71} < \pi < 3\frac{1}{7}$. Had he calculated an interval using the areas model, we could have applied this analysis to determine which method—areas or perimeters—gives the better approximation. The estimate for π using the areas method with $n = 96$ differs from the true value of π by 4.8052×10^{-7}, whereas the perimeters method for $n = 96$ differs by only 1.8022×10^{-7}, a 62% improvement over the areas approach. In fact, the perimeters approach is always more accurate than the areas approach. *Why?*

By using the improved perimeters method and comparing it to the error made with the areas method, we can see why the perimeters approach is more accurate. To find the error, we subtract π from our estimate:

$$\begin{aligned}\text{Error(perim)} &= \frac{2}{3}\left[n\sin\left(\frac{\pi}{n}\right)\right] + \frac{1}{3}\left[n\tan\left(\frac{\pi}{n}\right)\right] - \pi \\ &= \frac{2n}{3}\left[\frac{\pi}{1!n} - \frac{\pi^3}{3!n^3} + \frac{\pi^5}{5!n^5} - \frac{\pi^7}{7!n^7} + \cdots\right] \\ &\quad + \frac{n}{3}\left[\frac{\pi}{n} + \frac{\pi^3}{3n^3} + \frac{2\pi^5}{15n^5} + \frac{17\pi^7}{315n^7} + \cdots\right] - \pi\end{aligned}$$

Now we really get lucky. Not only do the linear terms in the two series sum to π—just as we designed them to do—but we get a bonus when the cubic terms also cancel each other precisely. We are left with the error in terms of order $O\left(\frac{1}{n^4}\right)$.

$$\text{Error(perim)} = \frac{\pi^5}{20n^4} + \frac{\pi^7}{56n^6} + \cdots$$

For comparison, the area method also gives an order $O\left(\frac{1}{n^4}\right)$ error term, specifically

$$\text{Error(area)} = \frac{4\pi^5}{45n^4} + \frac{2\pi^7}{63n^6} + \cdots$$

Comparing the ratio of the two errors relative to 1 will show that the perimeter estimate yields the smaller error

$$\frac{\text{Error(perim)}}{\text{Error(area)}} \approx \frac{\left(\frac{\pi^5}{20n^4}\right)}{\left(\frac{4\pi^5}{45n^4}\right)} = \frac{9}{16} = 0.5625.$$

But we needn't stop here! Once we know the size of the error in our formula, we can use it to remove most of the error. Based on the perimeter method, let's make *this* our new estimate:

$$\pi \approx \frac{2n}{3}\sin\left(\frac{\pi}{n}\right) + \frac{n}{3}\tan\left(\frac{\pi}{n}\right) - \frac{\pi^5}{20n^4}.$$

I can hear your objection already. Whereas before we did not need to know the value of π, we now need to know π to use this error formula. Well, not quite. Even a crude estimate of π on the right hand side will allow us to greatly reduce the error on the left. In fact, we can insert the value just computed in the current line of our table to improve the estimate. Our improved table is shown in TABLE 4.

TABLE 4: Removing the error

n-gon	Old estimate for π $\frac{2n}{3}\sin\left(\frac{\pi}{n}\right)+\frac{n}{3}\tan\left(\frac{\pi}{n}\right)$	Improved estimate for π $\frac{2n}{3}\sin\left(\frac{\pi}{n}\right)+\frac{n}{3}\tan\left(\frac{\pi}{n}\right)-\frac{\pi^5}{20n^4}$
6	**3.1**54 700 538 38	**3.142** 645 857 93
12	**3.142** 349 130 54	**3.141** 610 347 03
24	**3.141** 639 056 22	**3.141 592** 934 40
48	**3.141 595** 540 41	**3.141 592 657** 99
96	**3.141 592** 833 81	**3.141 592 653** 66

Notice that our adjustment has gained an extra three digits of accuracy without requiring any additional sides on the polygons. This is really no surprise, since the ratio of the new error term divided by the previous error term is

$$\frac{\left(\frac{\pi^7}{56n^6}\right)}{\left(\frac{\pi^5}{20n^4}\right)}=\frac{5\pi^2}{14n^2},$$

and for $n = 96$ this is about 0.000382. We should expect an additional three digits of accuracy. Perhaps that is enough, but we could now subtract the term of order $O\left(\frac{1}{n^6}\right)$ and gain even more accuracy.

Archimedes' argument that the constants of proportionality derived from the perimeters and areas of circles are equal to each other motivated intervals capturing but not giving a single approximated value for π. Remarkably, Taylor enables us to discover π's location in these intervals without knowing its actual value, creating a procedure for substantially improving the efficiency of Archimedes' intervals approach.

Just imagine if Archimedes had met Taylor . . .

REFERENCES

1. C. H. Edwards Jr., *The Historical Development of the Calculus*, Springer-Verlag, 1979.
2. Edwards and Penney, *Single Variable Calculus*, 5th ed., Prentice Hall, New Jersey, 1998.

Universality in Mathematical Modeling: A Comment on "Surprising Dynamics From a Simple Model"

HIDEYUKI SUZUKI
Institute of Industrial Science
The University of Tokyo
Tokyo 153–8505, Japan
hideyuki@sat.t.u-tokyo.ac.jp

KAZUYUKI AIHARA
Institute of Industrial Science
The University of Tokyo
Tokyo 153–8505, Japan
and
Aihara Complexity Modelling Project
ERATO, JST
Tokyo 102–0075, Japan
aihara@sat.t.u-tokyo.ac.jp

We enjoyed reading the article by Walsh in the December 2006 issue of this MAGAZINE [1], which introduces two simple models that exhibit rich nonlinear dynamics. It is especially notable that both the queueing model [2] and the cobweb model [3] are naturally derived from phenomena in the real world.

Mathematical modeling is powerful methodology in most fields in science, including physics, chemistry, biology, and engineering. If a phenomenon is described in the form of mathematical expressions, we can mathematically analyze and understand the phenomenon with various mathematical tools.

An important aspect of mathematical modeling is its universality. It is often the case that mathematical models for different phenomena in different fields share the same mathematical structure. In this case, results from a model in a certain field can be applied to other models in other fields through the common mathematical structure. An impressive example in nonlinear science is the Feigenbaum constant [4] in cascades of successive period-doubling bifurcations found in various phenomena.

In this Note, we introduce a few simple models with rich nonlinear dynamics in the fields of neuroscience and high voltage engineering. Surprisingly, the mathematical expressions for these models are identical or similar to those in [1], though they were proposed independently. This surprising coincidence can be considered as a good example illustrating the universality of mathematical modeling.

Nagumo-Sato neuron model

Neurons are the basic elements of nervous systems that realize information processing in the brain. Information in the nervous systems is transmitted between neurons mostly in the form of electrical pulses. Each neuron receives inputs from other neurons through synaptic connections, and when the amount of the inputs exceeds a certain threshold, the neuron emits an electrical pulse and outputs it to other neurons through synaptic connections.

The Nagumo-Sato neuron model [5] is a mathematical neuron model proposed on the basis of Caianiello's neuronic equations [6] to elucidate the mechanism of neuronal response. It can be considered as one of the simplest neuron models with refractoriness, which is the effect that discourages the neuron from firing twice in quick succession.

The Nagumo-Sato neuron model is a discrete-time model whose outputs are defined for time $t \in \{0, 1, 2, \dots\}$. Each output x_t at time t can take only the values 0 or 1, where 0 represents the resting state and 1 represents the firing state. The Nagumo-Sato neuron model can be described by the following equation:

$$x_{t+1} = H\left(A - \alpha \sum_{r=0}^{t} k^r x_{t-r} - \theta\right), \tag{1}$$

where A is the input strength which is assumed to be constant, $H(u)$ is the Heaviside step function ($H(u) = 1$ if $u \geq 0$, $H(u) = 0$ if $u < 0$), θ is the threshold value, and $\alpha > 0$ and $k \in [0, 1]$ are scaling and attenuation parameters for refractoriness, respectively. With Eq. (1), the output x_{t+1} at time $t + 1$ can be calculated from the history of the past firing sequence up to time t.

To investigate the dynamics of the model, we define the internal state y_{t+1} of the neuron at time $t + 1$ as follows:

$$y_{t+1} = A - \alpha \sum_{r=0}^{t} k^r x_{t-r} - \theta. \tag{2}$$

Then the output x_{t+1} at time $t + 1$ is simply determined from y_{t+1} by

$$x_{t+1} = H(y_{t+1}). \tag{3}$$

It is easy to check that y_{t+1} and y_t satisfy the following equation:

$$y_{t+1} = ky_t - \alpha H(y_t) + (1 - k)(A - \theta). \tag{4}$$

What should be emphasized here is that y_{t+1} is determined only by y_t. Therefore, Eq. (4) gives a one-dimensional dynamical system. Once the initial state y_0 is fixed, the orbit $y_0, y_1, y_2, \dots$ is iteratively determined by Eq. (4), and the sequence of the outputs $x_0, x_1, x_2, \dots$ are consequently determined by Eq. (3). Thus, Eq. (4) describes the dynamics of the Nagumo-Sato neuron model.

It is surprising that Eq. (4) is exactly equivalent to the function g in [**1**] (see Fig. 10 in [**1**]) that is obtained as an extreme case of the cobweb model. As noted for g in [**1**], the Nagumo-Sato neuron model has a stable periodic orbit for almost every input A, and the set of A's for which the model does not have any stable periodic orbit has fractal structure [**7**].

Chaotic neuron model

Although thresholding is an important process in real neurons, it is actually not as strict as the Heaviside step function. Rather, it should be described by continuous functions such as a sigmoidal function

$$\phi(u) = \frac{1}{1 + \exp(-u/\varepsilon)},$$

where the parameter $\varepsilon > 0$ represents strictness of thresholding. Thus, by replacing the Heaviside function $H(u)$ in the Nagumo-Sato neuron model (1) with a sigmoidal function $\phi(u)$, the chaotic neuron model [**8**] is obtained. The output x_{t+1} at time $t+1$ can be calculated from the history of past firing sequence up to time t as follows:

$$x_{t+1} = \phi\left(A - \alpha \sum_{r=0}^{t} k^r x_{t-r} - \theta\right). \tag{5}$$

We also define the internal state y_{t+1} by Eq. (2), exactly in the same way as for the Nagumo-Sato neuron model. Then the following equations hold:

$$x_{t+1} = \phi(y_{t+1}), \tag{6}$$

$$y_{t+1} = ky_t - \alpha\phi(y_t) + (1-k)(A-\theta). \tag{7}$$

Equation (7) provides a bimodal map with parameter regions where the dynamics are chaotic, as the name of the model implies [**8**]. The model given by Eqs. (6) and (7) can be extended to chaotic neural networks composed of chaotic neurons [**8, 9**], which produce rich nonlinear dynamics with engineering applicability [**10, 11, 12**]. Chaotic response was also observed in physiological experiments on real neurons [**13**].

Readers may notice that Eq. (7) is exactly identical to the equation for the cobweb model, as is Eq. (10) in [**1**]. The queueing model (cf. Eq. (7) in [**1**]) is also similar, but the coefficient k in the chaotic neuron model (7) is usually between 0 and 1, whereas the corresponding coefficient in the queueing model is precisely 1. The neuron model with $k = 1$ is physiologically implausible because it means that refractoriness lasts forever without any attenuation and the neuron eventually stops firing. However, despite this difference in k, they exhibit similar chaotic dynamics.

Partial-discharge model

The next model arises in high voltage engineering. Electrical discharges that only partially bridge the insulation between conductors are called partial discharges. Partial discharges in high-voltage systems are not directly disastrous, but they cause harmful effects such as energy loss and gradual degradation of the insulation, which may finally result in disastrous consequences. Thus analysis and diagnosis of partial discharges are important topics in high voltage engineering.

The three-capacitance model is the simplest model for partial-discharge phenomena. The model itself is old, proposed more than fifty years ago, but its complicated dynamics have been revealed only recently [**14**]. We omit detailed explanation of the model itself here, but its dynamics can be reduced to the class of double rotations [**15**]. A double rotation $f_{(\alpha,\beta,c)}: [0,1) \to [0,1)$ for $(\alpha,\beta,c) \in [0,1) \times [0,1) \times [0,1]$ is defined by

$$f_{(\alpha,\beta,c)}(x) = \begin{cases} \{x+\alpha\} & \text{if } x < c, \\ \{x+\beta\} & \text{if } x \geq c, \end{cases}$$

where $\{x\}$ denotes $x - \lfloor x \rfloor$. Double rotations can be considered as piecewise isometries, or more specifically, interval translation mappings [**16**].

Although double rotations are discontinuous and non-invertible in general, almost every double rotation can be reduced to a simple rotation, and the set of the parameter values for which a double rotation is not reducible to a rotation has fractal structure [**15, 17**]. Because of this self-similar structure in the parameter space of double rotations, the average discharge rate of the three-capacitance model as a function of the applied voltage is very complicated, resembling a devil's staircase, despite the simplicity of the model. This structure is quite similar to that of the Nagumo-Sato neuron model (and equivalently g in **[1]**).

The class of double rotations obviously includes simple rotations f_α in **[1]** that are derived as an extreme case of the queueing model. Though the coefficient for x is always 1 in both simple and double rotations, the dynamics of double rotations are nontrivial in general, because of the discontinuities.

Concluding remarks

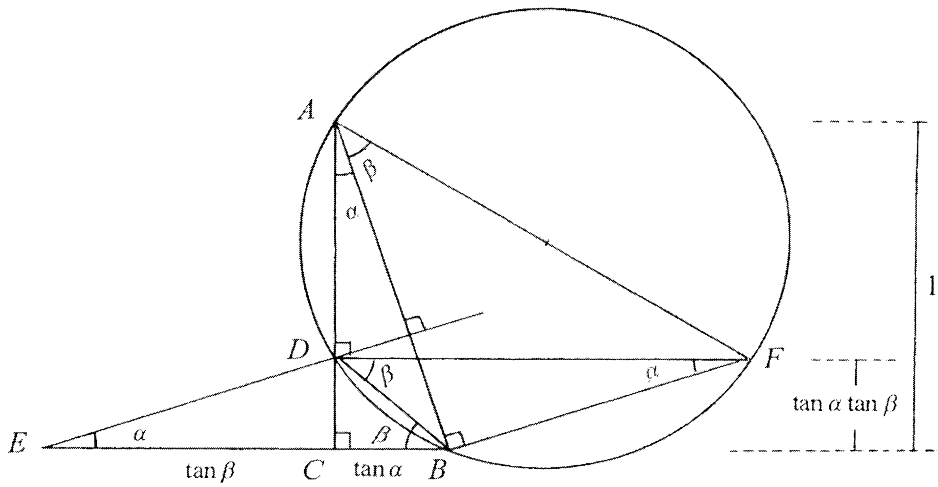
As pointed out in **[1]**, these simple models motivated by the real world show dynamics rich enough to be presented as examples of essential aspects of chaos theory. Moreover, these dynamics are closely related to contemporary research topics in mathematics such as the dynamics of piecewise isometries.

We also believe that the coincidences reported here illustrate the universality of mathematical modeling, for they give a strong impression that mathematical modeling can be considered as a "hub" that provides interdisciplinary connections among the various research fields which appear, at first glance, to be entirely different.

REFERENCES

1. J. A. Walsh, Surprising dynamics from a simple model, this MAGAZINE **79** (2006) 327–339.
2. G. Feichtinger, C. H. Hommes, and W. Herold, Chaos in a simple deterministic queueing system, *Math. Methods Oper. Res.* **40** (1994) 109–119.
3. C. H. Hommes, Dynamics of the cobweb model with adaptive expectations and nonlinear supply and demand, *J. Econom. Behav. Organ.* **24** (1994) 315–335.
4. M. J. Feigenbaum, The universal metric properties of nonlinear transformations, *J. Stat. Phys.* **21** (1979) 669–706.
5. J. Nagumo and S. Sato, On a response characteristics of a mathematical neuron model, *Kybernetik* **10** (1972) 155–164.
6. E. R. Caianiello, Outline of a theory of thought-processes and thinking machines, *J. Theoret. Biol.* **1** (1961) 204–235.
7. M. Hata, Dynamics of Caianiello's equation, *J. Math. Kyoto Univ.* **22** (1982) 155–173.
8. K. Aihara, T. Takabe, and M. Toyoda, Chaotic neural networks, *Phys. Lett. A* **144** (1990) 333–340.
9. M. Adachi and K. Aihara, Associative dynamics in a chaotic neural network, *Neural Netw.* **10** (1997) 83–98.
10. L. Chen and K. Aihara, Chaos and asymptotical stability in discrete-time neural networks, *Phys. D* **104** (1997), 286–325.
11. M. Hasegawa, T. Ikeguchi, and K. Aihara, Combination of chaotic neurodynamics with the 2-opt algorithm to solve traveling salesman problems, *Phys. Rev. Lett.* **79** (1997) 2344–2347.
12. L. Chen and K. Aihara, Global searching ability of chaotic neural networks, *IEEE Trans. Circuits Syst. I Regul. Pap.* **46** (1999) 974–993.
13. G. Matsumoto, K. Aihara, Y. Hanyu, N. Takahashi, S. Yoshizawa, and J. Nagumo, Chaos and phase locking in normal squid axons, *Phys. Lett. A* **123** (1987) 162–166.
14. H. Suzuki, K. Aihara, and T. Okamoto, Complex behaviour of a simple partial-discharge model, *Europhys. Lett.* **66** (2004) 28–34.
15. H. Suzuki, S. Ito, and K. Aihara, Double rotations, *Discrete Contin. Dyn. Syst.* **13** (2005) 515–532.
16. M. Boshernitzan and I. Kornfeld, Interval translation mappings, *Ergodic Theory Dynam. Systems* **15** (1995) 821–832.
17. H. Bruin and S. Troubetzkoy, The Gauss map on a class of interval translation mappings, *Israel J. Math.* **137** (2003) 125–148.

Proof Without Words: Tangent of the Sum



$$
\begin{aligned}
AC &= 1 \\
BC &= \tan\alpha & \Rightarrow \\
DC &= \tan\alpha\tan\beta & \Rightarrow \\
EC &= \tan\beta \\
DF &= EB & \Rightarrow
\end{aligned}
$$

$$\tan(\alpha+\beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta}$$

—Sidney H. Kung
Cupertino, CA 95014-2292
sidneykung@yahoo.com

An Elementary Proof of the Error Estimates in Simpson's Rule

D. D. HAI
Mississippi State University
Mississippi State, MS 39762

R. C. SMITH
Mississippi State University
Mississippi State, MS 39762

Recently, Cruz-Uribe and Neugebauer [**1**] gave an elementary proof of error estimates for the Trapezoidal rule, which based on integration by parts "backwards." However, they were unable to extend their method to the error estimates in Simpson's rule. In this note, we shall provide a new and elementary proof of error estimates in Simpson's rule for functions having four or three derivatives. Our approach also depends upon integration by parts and thus answers an open question posed in [**1**]. Note that in a more advanced paper [**2**], the authors derived error estimates in Trapezoidal and Simpson's rule for functions possessing just one or two derivatives. For other elementary proofs of error estimates in Simpson's rule, we refer to Talman [**4**] and the references therein.

To set our notation we let $f : [a, b] \to \mathbb{R}$ and for $n \in \mathbb{N}$, we put $x_j = a + \frac{j(b-a)}{2n}$ for $j = 0, 1, \dots, 2n$, and put

$$S_{2n} = \frac{b-a}{6n} \sum_{k=1}^{n} [f(x_{2k-2}) + 4f(x_{2k-1}) + f(x_{2k})].$$

If f is four times differentiable on $[a, b]$ and $K \geq 0$ is a constant such that $\left|f^{(4)}(x)\right| \leq K$ for all $x \in [a, b]$, then a standard result quoted in nearly every elementary calculus book is the error estimate

$$\left|\int_a^b f(x)\,dx - S_{2n}\right| \leq \frac{K(b-a)^5}{180(2n)^4}.$$

For $n > 1$, S_{2n} is called the composite Simpson's Rule in more advanced texts. We do the majority of our work in the situation where $n = 1$, $h > 0$, $[a, b] = [-h, h]$ and $S_2 = \frac{h}{3}[f(-h) + 4f(0) + f(h)]$. Put $E_2 = \int_{-h}^{h} f(x)\,dx - S_2$.

We are interested in estimates on E_2 for functions f that are at least three times differentiable on $[-h, h]$ and we do some preliminary calculations to write E_2 as an integral formula. In all subsequent integration by parts calculations throughout this paper, constants are chosen to make the polynomial factor have a zero at h.

Putting $g(x) = f(x) + f(-x)$ allows us to replace $\int_{-h}^{h} f(x)\,dx = \int_0^h g(x)\,dx$. Now integrate by parts, using $u = g(x)$, $dv = dx$, $du = g'(x)\,dx$, and $v = x - \frac{2h}{3}$ to get

$$\int_{-h}^{h} f(x)\,dx = \int_0^h g(x)\,dx = \frac{h}{3}g(h) + \frac{2h}{3}g(0) - \int_0^h \left(x - \frac{2h}{3}\right) g'(x)\,dx.$$

A second integration by parts, using $u = g'(x)$, $dv = \left(x - \frac{2h}{3}\right)dx$, $du = g''(x)\,dx$, and $v = \frac{x^2}{2} - \frac{2hx}{3} + \frac{h^2}{6}$ gives us

$$\int_{-h}^{h} f(x)\,dx = \frac{h}{3}[f(-h) + 4f(0) + f(h)] + \int_0^h \left(\frac{x^2}{2} - \frac{2hx}{3} + \frac{h^2}{6}\right) g''(x)\,dx.$$

Consequently,

$$E_2 = \int_0^h \left(\frac{x^2}{2} - \frac{2hx}{3} + \frac{h^2}{6}\right) g''(x)\,dx. \tag{1}$$

We start with the most familiar estimate:

THEOREM 1. *Let f be four times differentiable on $[-h, h]$ and suppose $K \geq 0$ is a constant such that $\left|f^{(4)}(x)\right| \leq K$ for all $x \in [a, b]$. Then $|E_2| \leq \frac{Kh^5}{90}$.*

Proof. From (1) and integrating by parts again, we obtain

$$E_2 = \int_0^h \left(\frac{x^2}{2} + \frac{2hx}{3} + \frac{h^2}{6}\right) g''(x)\,dx = -\int_0^h \left(\frac{x^3}{6} - \frac{hx^2}{3} + \frac{h^2x}{6}\right) g'''(x)\,dx. \tag{2}$$

Note that $g'''(0) = 0$, $|g^{(4)}(x)| \leq 2K$ for all $x \in [0, h]$, and $\frac{x^3}{6} - \frac{hx^2}{3} + \frac{h^2x}{6} = \frac{x(x-h)^2}{6} \geq 0$. By the Mean Value Theorem, for each $x \in (0, h]$, there is a point s_x in $(0, x)$ such that $g'''(x) = xg^{(4)}(s_x)$. Hence

$$|E_2| = \left|\int_0^h \left(\frac{x^3}{6} - \frac{hx^2}{3} + \frac{h^2x}{6}\right) xg^{(4)}(s_x)\, dx\right| \leq 2K \int_0^h \left(\frac{x^3}{6} - \frac{hx^2}{3} + \frac{h^2x}{6}\right) x\, dx$$
$$= \frac{Kh^5}{90}. \qquad \blacksquare$$

In both the statement and proof of the next result, we make use of real analysis. The particular notions and results from real analysis that we need are sometimes called the Fundamental Theorem of the Calculus for Lebesgue Integrals. All that we need can be found in pages 108–110 of [**3**]. We briefly summarize as follows: g is absolutely continuous on $[a, b]$ if and only if g' exists almost everywhere on $[a, b]$ and g' is Lebesgue integrable on $[a, b]$. In this case, $g(x) = g(a) + \int_a^x g'(t)\, dt$ for all $x \in [a, b]$. We let $L^1(a, b)$ denote the collection of Lebesgue integrable functions on $[a, b]$ and for $g \in L^1(a, b)$, we put $\|g\|_{L^1(a,b)} = \int_a^b |g(x)|\, dx$.

THEOREM 2. *Let f be three times differentiable on $[-h, h]$ and suppose f''' is absolutely continuous on $[-h, h]$. Then $|E_2| \leq \frac{h^4}{72}\|f^{(4)}\|_{L^1(-h,h)}$.*

Proof. Using (2) and integration by parts again give

$$E_2 = \int_0^h \left(\frac{x^4}{24} - \frac{hx^3}{9} + \frac{h^2x^2}{12} - \frac{h^4}{72}\right) g^{(4)}(x)\, dx. \tag{3}$$

Put $\phi(x) = \frac{x^4}{24} - \frac{hx^3}{9} + \frac{h^2x^2}{12} - \frac{h^4}{72}$ and note that $\phi(h) = 0$,

$$\phi'(x) = \frac{x^3}{6} - \frac{hx^2}{3} + \frac{h^2x}{6} = \frac{x(x-h)^2}{6} \geq 0 \quad \text{on } [0, h].$$

Thus $-\frac{h^4}{72} \leq \phi(x) \leq 0$, and so

$$|E_2| \leq \int_0^h |\phi(x)g^{(4)}(x)|\, dx \leq \frac{h^4}{72}\int_0^h |g^{(4)}(x)|\, dx = \frac{h^4}{72}\int_{-h}^h |f^{(4)}(x)|\, dx$$
$$= \frac{h^4}{72}\|f^{(4)}\|_{L^1(-h,h)}. \qquad \blacksquare$$

THEOREM 3. *Let f be three times differentiable on $[-h, h]$ and suppose there is a constant $M \geq 0$ such that $|f'''(x)| \leq M$ for all $x \in [-h, h]$. Then $|E_2| \leq \frac{Mh^4}{36}$.*

Proof. By the Mean Value Theorem, for each $x \in [0, h]$, there exists a point s_x between x and $h/3$ such that

$$g''(x) = g''(h/3) + (x - h/3)g'''(s_x).$$

Since $\int_0^h \left(\frac{x^2}{2} - \frac{2hx}{3} + \frac{h^2}{6}\right) dx = 0$ and $|g'''(x)| \leq 2M$ for all $x \in [0, h]$, it follows that

$$|E_2| = \left|\int_0^h \left(\frac{x^2}{2} - \frac{2hx}{3} + \frac{h^2}{6}\right) g''(x)\, dx\right|$$
$$= \left|\int_0^h \left(\frac{x^2}{2} - \frac{2hx}{3} + \frac{h^2}{6}\right)\left(x - \frac{h}{3}\right) g'''(s_x)\, dx\right|$$

$$= \frac{1}{18}\left|\int_0^h (3x-h)^2(x-h)g'''(s_x)\,dx\right|$$

$$\leq \frac{2M}{18}\int_0^h (3x-h)^2\,(h-x)\,dx = \frac{Mh^4}{36}. \qquad \blacksquare$$

We give the error estimates for the Composite Simpson's Rule as a corollary:

COROLLARY 4. *Let $f : [a, b] \to R$ be continuous, let S_{2n} be defined as above, let $n \in N$ and let $E_{2n} = \int_a^b f(x)\,dx - S_{2n}$. Then the following hold:*

(i) *if f is four times differentiable on $[a, b]$ and $K \geq 0$ is a constant such that $\left|f^{(4)}(x)\right| \leq K$ for all $x \in [a, b]$ then $|E_{2n}| \leq \frac{K(b-a)^5}{180(2n)^4}$;*

(ii) *if f is three times differentiable on $[a, b]$ and f''' is absolutely continuous on $[a, b]$ then $|E_{2n}| \leq \frac{(b-a)^4}{72(2n)^4}\left\|f^{(4)}\right\|_{L^1(a,b)}$;*

(iii) *if f is three times differentiable on $[a, b]$ and $M \geq 0$ is a constant such that $|f'''(x)| \leq M$ for all $x \in [a, b]$ then $|E_{2n}| \leq \frac{M(b-a)^4}{72(2n)^3}$.*

Proof. For the moment, fix $k \in \{1, \dots, n\}$. Let $h = \frac{b-a}{2n}$ and put $\tilde{g}(x) = f(x + x_{2k-1})$ for $-h \leq x \leq h$. Note that

$$\int_{x_{2k-2}}^{x_{2k}} f(x)\,dx - \frac{b-a}{6n}\left[f(x_{2k-2}) + 4f(x_{2k-1}) + f(x_{2k})\right]$$

$$= \int_{-h}^{h} \tilde{g}(x)\,dx - \frac{h}{3}\left[\tilde{g}(-h) + 4\tilde{g}(0) + \tilde{g}(h)\right].$$

If f is four times differentiable on $[a, b]$ and $K \geq 0$ is a constant such that $\left|f^{(4)}(x)\right| \leq K$ for all $x \in [a, b]$ then $\tilde{g}$ is four times differentiable on $[-h, h]$ and $\left|\tilde{g}^{(4)}(x)\right| = \left|f^{(4)}(x + x_{2k-1})\right| \leq K$ for all $x \in [-h, h]$. By Theorem 1,

$$\left|\int_{-h}^{h} \tilde{g}(x)\,dx - \frac{h}{3}\left[\tilde{g}(-h) + 4\tilde{g}(0) + \tilde{g}(h)\right]\right| \leq \frac{Kh^5}{90} = \frac{K(b-a)^5}{90(2n)^5}.$$

If f is three times differentiable on $[a, b]$ and f''' is absolutely continuous on $[a, b]$ then $\tilde{g}$ is three times differentiable on $[-h, h]$ and $\tilde{g}'''$ is absolutely continuous on $[-h, h]$. Moreover,

$$\left\|\tilde{g}^{(4)}\right\|_{L^1(-h,h)} = \int_{-h}^{h} \left|\tilde{g}^{(4)}(x)\right|dx = \int_{x_{2k-2}}^{x_{2k}} \left|f^{(4)}(x)\right|dx,$$

and so by Theorem 2,

$$\left|\int_{-h}^{h} \tilde{g}(x)\,dx - \frac{h}{3}\left[\tilde{g}(-h) + 4\tilde{g}(0) + \tilde{g}(h)\right]\right|$$

$$\leq \frac{h^4}{72}\left\|\tilde{g}^{(4)}\right\|_{L^1(-h,h)} = \frac{(b-a)^4}{72(2n)^4}\int_{x_{2k-2}}^{x_{2k}} \left|f^{(4)}(x)\right|dx.$$

If f is three times differentiable on $[a, b]$ and $M \geq 0$ is a constant such that $|f'''(x)| \leq M$ for all $x \in [a, b]$ then $\tilde{g}$ is three times differentiable on $[-h, h]$ and $|\tilde{g}'''(x)| \leq M$ for all $x \in [-h, h]$ so by Theorem 3,

$$\left|\int_{-h}^{h} \tilde{g}(x)\,dx - \frac{h}{3}\left[\tilde{g}(-h) + 4\tilde{g}(0) + \tilde{g}(h)\right]\right| \leq \frac{Mh^4}{36} = \frac{M(b-a)^4}{36(2n)^4}.$$

The estimates in (i) and (iii) are now a matter of summing over k and using the triangle inequality:

In (i) we have

$$\begin{aligned}|E_{2n}| &= \left|\sum_{k=1}^{n}\left(\int_{x_{2k-2}}^{x_{2k}} f(x)\,dx - \frac{b-a}{6n}\left[f(x_{2k-2}) + 4f(x_{2k-1}) + f(x_{2k})\right]\right)\right| \\ &\le \sum_{k=1}^{n}\left|\int_{x_{2k-2}}^{x_{2k}} f(x)\,dx - \frac{b-a}{6n}\left[f(x_{2k-2}) + 4f(x_{2k-1}) + f(x_{2k})\right]\right| \\ &\le n\frac{K(b-a)^5}{90(2n)^5} = \frac{K(b-a)^5}{180(2n)^4},\end{aligned}$$

while in (iii) we have

$$|E_{2n}| \le n\frac{M(b-a)^4}{36(2n)^4} = \frac{M(b-a)^4}{72(2n)^3}.$$

The estimate in (ii) is also obtained by summing over k but we make use of

$$\|f^{(4)}\|_{L^1(a,b)} = \int_a^b |f^{(4)}(x)|\,dx = \sum_{k=1}^{n}\int_{x_{2k-2}}^{x_{2k}} |f^{(4)}(x)|\,dx$$

to see that

$$|E_{2n}| \le \sum_{k=1}^{n}\frac{(b-a)^4}{72(2n)^4}\int_{x_{2k-2}}^{x_{2k}} |f^{(4)}(x)|\,dx = \frac{(b-a)^4}{72(2n)^4}\|f^{(4)}\|_{L^1(a,b)}.$$

Remarks. (i) In Theorem 1, if we assume $f^{(4)}$ is continuous on $[-h, h]$, then we can avoid the use of the Mean Value Theorem and simply use the formula (3) to obtain

$$|E_2| \le -2K\int_0^h\left(\frac{x^4}{24} - \frac{hx^3}{9} + \frac{h^2x^2}{12} - \frac{h^4}{72}\right)dx = \frac{Kh^5}{90}.$$

A similar remark can be made about Theorem 3.

(ii) An even better approach is to use the Lebesgue integral instead of the Riemann integral. In this situation, it is enough to assume $f^{(3)}$ is absolutely continuous and $f^{(4)}$ is essentially bounded. Moreover, K can be replaced by the essential supremum of $|f^{(4)}|$ and we don't need a uniform bound on $f^{(4)}$. The major disadvantage of this approach (and our Theorem 2) is that it requires more analysis knowledge than we find with typical undergraduates.

(iii) We note that all of our error estimates are sharp. The central result in [**4**] gives an accessible argument that Theorem 1 is sharp. To see that Theorem 2 is sharp, consider the situation with $0 < \varepsilon < h$ and

$$f_\varepsilon(x) = \begin{cases} 0 & \text{for } x < -\varepsilon, \\ \frac{\varepsilon^3}{72} + \frac{\varepsilon^2 x}{15} + \frac{\varepsilon x^2}{8} + \frac{x^3}{9} + \frac{x^4}{24\varepsilon} - \frac{x^6}{360\varepsilon^3} & \text{for } |x| \le \varepsilon, \\ \frac{2\varepsilon^2 x}{15} + \frac{2x^3}{9} & \text{for } x > \varepsilon. \end{cases}$$

It is tedious but straightforward to verify that f_ε is four times continuously differentiable on $[-h, h]$ and $\frac{h^4}{72}\int_{-h}^{h}|f_\varepsilon^{(4)}(x)|\,dx = \frac{h^4}{54}$ and $\frac{h^4}{54} - |E_2| \le O(\varepsilon^2)$.

Finally, to see that Theorem 3 is sharp, we again consider the case when $0 < \varepsilon < h$, but this time we put

$$f_\varepsilon(x) = \begin{cases} \frac{x^5}{30\varepsilon^2} & \text{for } 0 \le x \le \varepsilon, \\ \frac{x^3}{3} - \frac{2\varepsilon x^2}{3} + \frac{\varepsilon^2 x}{2} - \frac{2\varepsilon^3}{15} & \text{for } \varepsilon < x \le h, \end{cases}$$

and extend f_ε to be an even function. Another straightforward but tedious argument shows that f_ε is three times continuously differentiable and $\left|f_\varepsilon'''(x)\right| \le 2$ for all $x \in [-h, h]$. In this case $M = 2$ and our estimate in Theorem 3 becomes $\frac{h^4}{18}$; a direct calculation shows $\frac{h^4}{18} - |E_2| \le O(\varepsilon^2)$. Note that in [**4**], Talman shows that $|E_2| \le \frac{2Mh^4}{9}$. This is similar in form to our result but not sharp as Talman himself points out.

Acknowledgements. The authors wish to thank the referees for pointing out some errors in the examples in an earlier version of this paper and for their suggestions for improving our exposition.

REFERENCES

1. D. Cruz-Uribe and C. J. Neugebauer, An elementary proof of error estimates for the trapezoidal rule, this MAGAZINE **76** (2003) 303–306.
2. D. Cruz-Uribe and C. J. Neugebauer, Sharp error bounds for the trapezoidal rule and Simpson's rule, *J. Ineq. Pure Appl. Math* **3**(4) (2002) article 49.
3. H. L. Royden, *Real Analysis*, 3rd Ed., Macmillan, New York, 1988.
4. L. A. Talman, Simpson's rule is exact for quintics, *Amer. Math. Monthly* **113** (2006) 144–155.

The Class of Heron Triangles

RAYMOND A. BEAUREGARD
University of Rhode Island
Kingston, RI 02881
beau@math.uri.edu

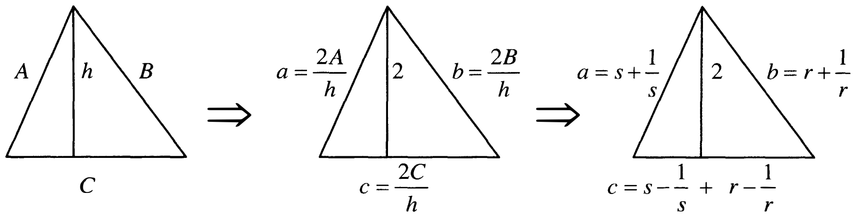
A Heron triangle is one with rational sides and rational area. A glance at sources ranging from Dickson's wonderful compilation [**1**] to the modern Wolfram website [**2**] indicates that such triangles have not only been but will continue to be a fascination. If a Heron triangle is scaled up using the lowest common multiple of the denominators of the sidelengths, then the triangle is similar to one having integer sides and integer area. Three integer parameters m, n, k can then be given that generate all such Heron triangles: the sides have the form (see [**2**])

$$n(m^2 + k^2),\ m(n^2 + k^2),\ (m + n)(mn - k^2).$$

In this note we start with a Heron triangle and scale it (usually down) to produce a Heron triangle with altitude 2. We are then able to obtain two rational parameters that generate this family of Heron triangles.

THEOREM. *After scaling, every Heron triangle has the form of the right-most triangle in the figure below, for rationals r and s with* $r, s > 1$.

Proof. Any triangle can be oriented so that it has the configuration of the first triangle in the figure. If we assume that it is a Heron triangle and if we multiply its sides A, B, C and its (rational) altitude h by the scaling factor $2/h$, we obtain the second triangle in the figure with rational sides a, b, c and altitude 2.



Since $a, b > 2$ there are real numbers $r, s > 1$ such that $a = s + 1/s$ and $b = r + 1/r$. One simply solves the equations $s^2 - as + 1 = 0$ and $r^2 - br + 1 = 0$, obtaining

$$s = \frac{a + \sqrt{a^2 - 4}}{2}, \quad r = \frac{b + \sqrt{b^2 - 4}}{2}.$$

Now c is sum of the bases of the two right triangles in the figure (third triangle); using the Pythagorean theorem twice we see that $c = s - 1/s + r - 1/r$. It remains to show that both r and s are rational. But this follows from the equations

$$\begin{aligned}
\frac{a+b+c}{2} &= r + s, \\
\frac{a+b-c}{2} &= \frac{1}{r} + \frac{1}{s}, \\
\frac{a-b+c}{2} &= s - \frac{1}{r}, \\
\frac{-a+b+c}{2} &= r - \frac{1}{s}.
\end{aligned}$$

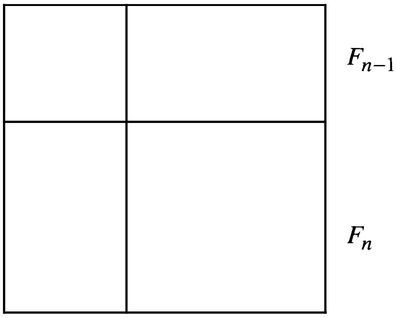
Dividing the first equation by the second shows that rs is rational; the third equation then shows that r is rational, and s follows suit because of the fourth equation. ∎

The interested reader is invited to show how the (r, s) parameters yield the (m, n, k) parameters.

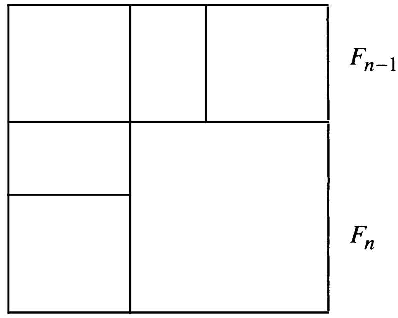
REFERENCES

1. Leonard E. Dickson, *History of the Theory of Numbers*, Vol. II, Chelsea, New York, 1952.
2. Eric W. Weisstein, "Heronian Triangle," Math World: A Wolfram Websource, `http://mathworld.wolfram.com/HeronianTriangle.html`

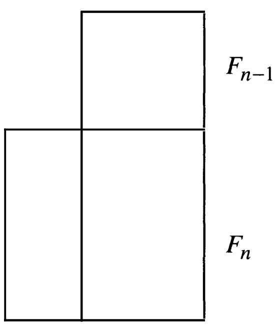
Proof Without Words: Fibonacci Tiles



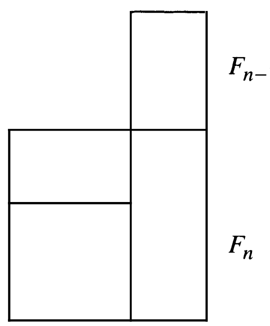
$$
\begin{aligned}
F_{n+1}^2 &= 2F_{n+1}F_n - F_n^2 + F_{n-1}^2 \\
&= 2F_{n+1}F_{n-1} + F_n^2 - F_{n-1}^2 \\
&= 2F_n F_{n-1} + F_n^2 + F_{n-1}^2 \\
&= F_{n+1}F_n + F_n F_{n-1} + F_{n-1}^2 \\
&= F_{n+1}F_{n-1} + F_n^2 + F_n F_{n-1}
\end{aligned}
$$



$$F_{n+1}^2 = F_n^2 + 3F_{n-1}^2 + 2F_{n-1}F_{n-2}$$



$$F_n^2 = F_{n+1}F_{n-1} + F_nF_{n-2} - F_{n-1}^2$$



$$F_n = F_{n+1}F_{n-2} + F_{n-1}^2$$

F_n denotes the nth Fibonacci number, where $F_{n+1} = F_n + F_{n-1}$, $F_0 = 0$, $F_1 = 1$. Obvious assumptions concerning the least value of n in each identity should be made as required. Further visual proofs of Fibonacci identities may be found in:

A. Brousseau, Fibonacci numbers and geometry, *Fibonacci Quarterly* **10** (1972) 308–318.

M. Bicknell-Johnson and D. DeTemple, Visualizing golden ratio sums with tiling patterns, *Fibonacci Quarterly* **33** (1995) 298–303.

—Richard L. Ollerton
School of Computing & Mathematics
University of Western Sydney
Kingswood DC1797, Australia
r.ollerton@uws.edu.au

PROBLEMS

ELGIN H. JOHNSTON, *Editor*
Iowa State University

PROPOSALS

To be considered for publication, solutions should be received by March 1, 2009.

2001. *Proposed by José H. Nieto, Universidad del Zulia, Maracaibo, Venezuela.*

A chess club has n members. Each member of the club has played against all but k of the other members. The club decides to hold a tournament in which each member plays exactly one game against those he/she has not played before. The tournament is played in rounds, with each player playing at most one game each round. Each round is scheduled by randomly selecting pairs who have not previously played against each other (and who are not already scheduled for the round) until no more such pairs are available for the round. Determine the maximum possible number of rounds for such a tournament. (For example, if the club has six members A, B, C, D, E, and F and the pairs that have never played against each other are AB, AC, AD, BC, BE, CF, DE, DF, and EF, then the tournament could consist of the following rounds:

$$1: AB, CF, DE, \qquad 2: BC, DF \qquad 3: AC, EF \qquad 4: AD, BE.)$$

2002. *Proposed by Dorin Marghidanu, Colegiul National "A. I. Cuza," Corabia, Romania.*

Let $a_1, a_2, \ldots, a_n$ be positive real numbers. Prove that

$$\frac{a_1^2}{a_1+a_2} + \frac{a_2^2}{a_2+a_3} + \cdots + \frac{a_{n-1}^2}{a_{n-1}+a_n} + \frac{a_n^2}{a_n+a_1} \geq \frac{a_1+a_2+\cdots+a_n}{2}.$$

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, written or electronic, should include **on each page** the reader's name, full address, and an e-mail address and/or FAX number.

2003. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

Let $(X, \langle\,\rangle)$ be a real inner product space, and let

$$B = \{x \in X : \|x| \leq 1\}$$

be the unit ball in X, where $\|x\| = \sqrt{\langle x, x\rangle}$. Let $f : B \to B$ be a function satisfying $\|f(x) - f(y)\| \leq \|x - y\|$ for all $x, y \in B$. Prove that the set of fixed points of f is convex.

2004. *Proposed by Jody M. Lockhart and William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

Let A be an $n \times n$ matrix over the finite field F_q of q elements, and assume that A has multiplicative order $\mathrm{ord}(A) = q^n - 1$. Prove or give a counterexample to the following statement:

> A is a cyclic generator for F_{q^n}, that is, $\{0, A, A^2, \ldots, A^{q^n-1}\}$ is the finite field F_{q^n}.

2005. *Proposed by Ovidiu Furdui, Campia Turzii, Cluj, Romania.*

Let $f : [0, \infty) \to (0, \infty)$ be an increasing, differentiable function with continuous derivative, and let k be a nonnegative integer. Prove that

$$\int_0^\infty \frac{x^k}{f(x)}\,dx$$

converges if and only if

$$\int_0^\infty \frac{x^k}{f(x) + f'(x)}\,dx$$

converges.

Quickies

Answers to the Quickies are on page 308.

Q983. *Proposed by Michel Bataille, Rouen, France.*

Let a, b, c be real numbers with $a + b + c = 0$ and $a^2 + b^2 + c^2 = 1$. Show that

$$a^4 + b^4 + c^4 - (a^3 + b^3 + c^3)^2 \geq \frac{1}{3},$$

and determine conditions under which equality holds.

Q984. *Proposed by Ovidiu Furdui, Campia Turzii, Cluj, Romania.*

Let k and p be nonnegative integers. Show that

$$\int_{-1}^{1} \frac{x^{2p}}{1 + x^{2k+1} + \sqrt{1 + x^{4k+2}}}\,dx = \frac{1}{2p+1}.$$

Solutions

Increasing zeroes **October 2007**

1776. *Proposed by Leon Gerber, St. John's University, Jamaica, NY.*

Let n be an odd integer, let $f_n(x) = (1+x)^n - (1+nx)$, and let x_n be the unique negative solution to $f_n(x) = 0$. It is easy to show that f_n has a positive relative maximum at $x = -2$. Prove that the sequence $\{f_n(-2)\}$ is increasing, and that $\lim_{n\to\infty} x_n = -2$.

Solution by Michael Woltermann, Washington and Jefferson College, Washington, PA. Let $n = 2k+1$. Then $f_n(-2) = 4k$, so the sequence $\{f_n(-2)\}$ is increasing. For sufficiently large n,

$$f_n\left(-2 - \frac{1}{\sqrt{n}}\right) = -\left(\left(1 + \frac{1}{\sqrt{n}}\right)^{\sqrt{n}}\right)^{\sqrt{n}} + 2n + \sqrt{n} - 1 < 0.$$

By the Intermediate Value Theorem, it follows that $-2 - \frac{1}{\sqrt{n}} < x_n < -2$, and hence that $\lim_{n\to\infty} x_n = -2$.

Also solved by Michael Andreoli, Brian D. Beasley, Paul Bracken and P. Lindberg, Brian Bradie, Bruce S. Burdick, Robert Calcaterra, John Ernest Chavez, Jamala Cooper, Chip Curtis, Knut Dale (Norway), Paul Deiermann, The Fisher Problem Solving Group, Dmitry Fleischman, Richard Gray, Dionissius Gressis and Petros Gressis, Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Albania), Elias Lampakis (Greece), Leo Livshutz, José H. Nieto (Venezuela), Kees Onneweer, Nicholas C. Singer, Albert Stadler (Switzerland), David Stone and John Hawkins, Marian Tetiva (Romania), Nora Thornber, Bob Tomper, Michael Vowe (Switzerland), Xiaoshen Wang, Kevin Weis, and the proposer. There was one incorrrect submission.

Polygon equations **October 2007**

1777. *Proposed by Richard A. Jacobson, Houghton College, Houghton, NY.*

The graph of the relation $|x+y| + |x-y| = 2$ is a square of side length 2. Find positive integer n and real constants a_k, b_k, c_k, $1 \le k \le n$ such that the graph of the relation $\sum_{k=1}^{n} |a_k x + b_k y + c_k| = 2$ is

a. a regular hexagon of side length 2.

b. a regular dodecagon of side length 2.

Solution by Victor Y. Kutsenok, University of St. Francis, Fort Wayne, IN.

a. Consider the regular hexagon of side two, centered at the origin, and with one vertex at (2, 0). Because the polygon is symmetric about the two axes and has ± 1 and ± 2 as the x-coordinates of its vertices, we consider an equation of the form

$$a|x+1| + b|y| + a|x-1| = 2. \tag{1}$$

Substituting $(0, \sqrt{3})$ and $(2, 0)$ into (1) we find $a = \frac{1}{2}$ and $b = \frac{1}{\sqrt{3}}$. When these values are substituted into (1), it is easy to check that the graph of the resulting equation is a polygon with vertices $(\pm 2, 0)$, $(\pm 1, \sqrt{3})$, and $(\pm 1, -\sqrt{3})$, so is a hexagon as desired.

b. Now consider the regular dodecagon of side two, centered at the origin, and with four vertices on the coordinate axes. Then the vertices in the first quadrant are

$$(0, r), \quad \left(\frac{r}{2}, \frac{r\sqrt{3}}{2}\right), \quad \left(\frac{r\sqrt{3}}{2}, \frac{r}{2}\right), \quad \text{and} \quad (r, 0), \tag{2}$$

where $r = \sqrt{6} + \sqrt{2}$. Because of the symmetry of the figure and the position of the x-coordinates of the vertices, we consider an equation of the form

$$a|x| + b\left|x + \frac{r}{2}\right| + b\left|x - \frac{r}{2}\right| + c\left|x + \frac{r\sqrt{3}}{2}\right| + c\left|x - \frac{r\sqrt{3}}{2}\right| + d|y| = 2. \quad (3)$$

Substituting the points in (2) into (3) we find

$$a = \frac{\sqrt{2}}{2}(11\sqrt{3} - 19), \quad b = \frac{\sqrt{2}}{4}(14 - 8\sqrt{3}),$$

$$c = \frac{\sqrt{2}}{4}(4 - 2\sqrt{3}), \quad d = \frac{\sqrt{2}}{2}(3\sqrt{3} - 5).$$

Using these values for the coefficients in (3), we find the graph of the resulting equation is a regular dodecagon as desired.

Also solved by Brian Bradie, Bruce S. Burdick, Chip Curtis, Jim Delany, M. J. Engelfield (Australia), Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Albania), Elias Lampakis (Greece), Kim McInturff, José H. Nieto (Venesuela), Samuel Otten, and the proposer.

Determinant counting **October 2007**

1778. *Proposed by Jody M. Lockhart and William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

Let q be a positive integer power of a prime and let F_q denote the finite field of q elements. For each positive integer n and each $\gamma \in F_q$, find the number of $n \times n$ matrices over F_q with determinant γ.

Solution by Chip Curtis, Missouri Southern State University, Joplin, MO.
Let $M_q(n)$ be the set of $n \times n$ matrices over F_q, let $N(n, q, \gamma)$ denote the number of elements of $M_q(n)$ with determinant γ, and let 1_q denote the multiplicative identity in F_q. If $A \in M_q(n)$ with $\det A = 1_q$, and $f \in F_q$ with $f \neq 0$, then, multiplying the first row of A by f gives a matrix $A' \in M_q(n)$ with $\det A' = f$. Because $f \neq 0$ has a multiplicative inverse in F_q, it follows that the map $A \to A'$ is a bijection between the set of matrices with determinant 1 and the set of matrices with determinant f. Thus $N(n, q, f) = N(n, q, 1_q)$ for all nonzero $f \in F_q$.

We now count the number of invertible matrices A in $M_q(n)$. The first row of A can be anything except for the zero row, so there are $q^n - 1$ possibilities for this row. The second row can be anything that is not a multiple of the first row, so there are $q^n - q$ choices for this row. The third row can be anything that is not a linear combination of the first two rows, giving $q^n - q^2$ possibilities for this row. Continuing in this manner, we find that the number of invertible matrices in $M_q(n)$ is $\prod_{k=1}^{n}(q^n - q^{k-1})$. It then follows that

$$N(n, q, 0) = q^{n^2} - \prod_{k=1}^{n}(q^n - q^{k-1}),$$

and that for nonzero $f \in F_q$,

$$N(n, q, f) = \frac{1}{q-1}\prod_{k=1}^{n}(q^n - q^{k-1}).$$

Also solved by Paul Budney, Bruce S. Burdick, Robert Calcaterra, Jim Delany, Toni Ernvall, Eugene A. Herman, Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Albania), Elias Lampakis (Greece), The Missouri State University Problem Solving Group, José H. Nieto (Venezuela), Samuel Otten and Chris Zin, Éric Pité (France), Marian Tetiva (Romania), Bob Tomper, and the proposer.

Walking the lines **October 2007**

1779. *Proposed by Will Gosnell, Amherst, MA., and Herb Bailey, Rose Hulman Institute of Technology, Terre Haute, IN.*

Let ABC be a triangle with $BC = a$, $CA = b$, and $AB = c$, let $\theta = \angle ACB$, and let $k = c/b$. Alice walks from C to A at a constant speed, and Bob walks from B to A, also at constant speed. The sum of the two travel speeds is numerically equal to a, and for each the travel time is numerically equal to $k^{3/2}$. For which values of k can a value of θ be found so that the given conditions can be satisfied? (This problem generalizes the problem with $\theta = 90°$ posed by Will Gosnell in the February 2005 issue of *Math Horizons.*)

Solution by Nicholas C. Singer, Annandale, VA.
Let Alice's and Bob's speeds be v_A and v_B, respectively. Then

$$v_A + v_B = a, \quad k^{3/2} = \frac{b}{v_A} = \frac{c}{v_B}, \quad \text{and} \quad c^2 = a^2 + b^2 - 2ab\cos\theta.$$

Because $c = bk$, we have $k^{3/2}a = b + c = b(k+1)$ and $b^2k^2 = a^2 + b^2 - 2ab\cos\theta$. Thus $b = \frac{k^{3/2}a}{k+1}$, so since $k > 0$ and $a > 0$,

$$\frac{k^5}{(k+1)^2} = 1 + \frac{k^3}{(k+1)^2} - 2\frac{k^{3/2}}{k+1}\cos\theta.$$

Hence,

$$k^5 = (k+1)^2 + k^3 - 2(k+1)k^{3/2}\cos\theta,$$

and therefore

$$\cos\theta = \frac{(k+1)^2 + k^3 - k^5}{2(k+1)k^{3/2}} = \frac{1 + k + k^3 - k^4}{2k^{3/2}}.$$

This is feasible if and only if $\cos^2\theta \le 1$. Thus we must have $(1 + k + k^3 - k^4)^2 \le 4k^3$, that is,

$$(k-1)(k^2 + k + 1)(k^5 - 2k^4 + k^3 - k^2 - 2k - 1) \le 0.$$

Because $k > 0$, this condition amounts to

$$(k-1)(k^5 - 2k^4 + k^3 - k^2 - 2k - 1) \le 0.$$

This holds for $1 \le k \le r$, where r is the real zero of $k^5 - 2k^4 + k^3 - k^2 - 2k - 1$. We have $r \approx 2.04235$.

Also solved by Michael Andreoli, Saadia Borders, Brian Bradie, Bruce S. Burdick, John Christopher, Chip Curtis, Knut Dale (Norway), Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Albania), Victor Y. Kutsenok, Hannah Savage, Michael Vowe (Switzerland), and the proposer.

Counting the odd digits **October 2007**

1780. *Proposed by Yiu Tung Poon, Iowa State University, Ames, IA.*

For positive integer k, define $\mathrm{odd}(k)$ to be the number of odd digits in the (base-ten) expansion of 2^k. Prove that

$$\sum_{k=1}^{\infty} \frac{\mathrm{odd}(k)}{2^k} = \frac{1}{9}.$$

Solution by The George Washington University Problems Group, Washington DC.
For positive integers k and j define

$$f(k, j) = \begin{cases} 1 & \text{if } \left\lfloor \dfrac{2^k}{10^j} \right\rfloor \text{ is an odd integer,} \\ 0 & \text{otherwise.} \end{cases}$$

Thus also $f(k, j)$ is 1 if and only if the jth decimal digit of 2^k, counting from the right, is odd. The key observation is that in addition, $f(k, j)$ is 1 if and only if the kth binary digit of $\frac{1}{10^j}$, counting right from the decimal point, is odd, that is, equals 1. Hence

$$\sum_{k=1}^{\infty} \frac{\mathrm{odd}(k)}{2^k} = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \frac{f(k, j)}{2^k} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \frac{f(k, j)}{2^k} = \sum_{j=1}^{\infty} \frac{1}{10^j} = \frac{1}{9}.$$

Note. Several readers noted that a generalization of this problem appears on page 15 of *Experimentation in Mathematics: Computational Paths to Discovery*, by J. Borwein, D. Bailey, and R. Girgensohn.

Also solved Bruce S. Burdick, John Christopher, G.R.A.20 Problems Solving Group (Italy), João Guerrerio (Portugal), Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Albania), Elias Lampakis (Greece), Peter W. Lindstrom, Ryan Muller, José H. Nieto (Venezuela), Northwestern University Math Problem Solving Group, Nicholas C. Singer, Albert Stadler (Switzerland), G. Wildenberg, and the proposer.

Answers

Solutions to the Quickies from page 304.

A983. Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{pmatrix} \quad \text{and} \quad s_k = a^k + b^k + c^k, \quad k = 3, 4.$$

The desired inequality follows from

$$0 \le (\det A)^2 = \det(AA^T) = \det \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & s_3 \\ 1 & s_3 & s_4 \end{pmatrix} = 3s_4 - 3s_3^2 - 1.$$

Because $\det(A) = (b-a)(c-a)(c-b)$, equality holds if and only if two of the numbers a, b, c are equal. Given the conditions on a, b, c, we see that equality holds if and only if two of the numbers are equal to $\frac{\sqrt{6}}{6}$ or two are equal to $-\frac{\sqrt{6}}{6}$.

A984. Let A denote the value of the integral. Making the substitution $x = -y$ we obtain

$$A = \int_{-1}^{1} \frac{y^{2p}}{1 - y^{2k+1} + \sqrt{1 + y^{4k+2}}}\, dy.$$

Thus

$$\begin{aligned} A = \frac{A + A}{2} &= \frac{1}{2} \int_{-1}^{1} \frac{2x^{2p}\left(1 + \sqrt{1 + x^{4k+2}}\right)}{\left(1 - x^{2k+1} + \sqrt{1 + x^{4k+2}}\right)\left(1 + x^{2k+1} + \sqrt{1 + x^{4k+2}}\right)}\, dx \\ &= \frac{1}{2} \int_{-1}^{1} x^{2p} dx = \frac{1}{2p + 1}. \end{aligned}$$

You Say "Viete," I Say "Vieta"

Rick Kreminski writes that readers might enjoy two references [**1, 2**] he recently found. They include improvements to Archimedes's approach to pi that use Richardson's extrapolation, somewhat like in his recent paper [**3**] on accelerating Vieta's formula (although they take a different approach, and do not supply error bounds nor high-precision numerics). He also points out that while "Vieta" is the spelling used in various articles from the Monthly and this MAGAZINE, the name is also often spelled "Viete" (sometimes with an accent). For a variety of spellings, see the references in [**4**] (which includes a Latin version, "VIETÆ").

1. G. Miel, Of calculations past and present: the Archimedean algorithm, *American Mathematical Monthly* **90** (1983) 17–35.
2. S. Schonefeld, Some examples illustrating Richardson's improvement, *College Mathematics Journal* **21** (1990) 314–236.
3. R. Kreminski, Pi to thousands of digits from Vieta's formula, this MAGAZINE **81** (2008) 201–207.
4. L. Berggren, J. Borwein, and P. Borwein, *Pi: A Source Book*, 3rd ed., Springer-Verlag, 2004.

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles, books, and other materials 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.

Kriz, Igor, and Paul Siegel, Simple groups at play, *Scientific American* (July 2008) 84–89; `http://www.sciam.com/article.cfm?id=simple-groups-at-play`.

Rubik's Cube and most other permutation puzzles yield to the same solution strategies (commutators, macros, etc.). The authors offer, as new vehicles for developing intuition about groups, three new puzzles (two to play online, one to download) whose solutions require different techniques. The puzzles are constructed from sporadic simple groups, whereas the usual permutation puzzles are based on the symmetric and alternating groups S_n and A_n. The three sporadic simple groups are the Mathieu groups M_{12} and M_{24} and the Conway group Co_0 (also known as Dotto). The distinguishing feature of the new puzzles is that each group is a "small" subgroup of the corresponding symmetric group, so that many arrangements are inaccessible—just as the odd permutations are unreachable in the classic 15-puzzle. (The authors are a professor and an undergraduate, and the work originated in an NSF-sponsored Research Experience for Undergraduates project.)

Fountain, Henry, A problem of bubbles frames an Olympic design, *New York Times* (5 August 2008) D4 (National ed.), F4 (New York ed.); `http://www.nytimes.com/2008/08/05/sports/olympics/05swim.html` (with correction appended 8 August 2008). Rehmeyer, Julie, A building of bubbles, *Science News* (19 July 2008); `http://www.sciencenews.org/view/generic/id/34283/title/A_building_of_bubbles`. Weaire, D., and R. Phelan, a counterexample to Kelvin's conjecture on minimal surfaces, *Philosophical Magazine Letters* 69 (2) (February 1994) 107–110; reprinted on pp. 47–52 of Weaire, Denis (ed.), *The Kelvin Problem: Foam Structures of Minimal Surface Area*, CRC Press, 1997. ISBN 0–74840632–8.

I write this just before the Olympics open in Beijing but after glimpsing on TV the mesh appearance of the Beijing National Aquatics Center (the "Water Cube"). It's too bad that you will read this after the Olympics, because the building has an interesting mathematical feature that you could point out to others. The architect, eager to have the structure reflect a "connection with water," came across bubbles and foams and eventually found the indicated article by Weaire and Phelan. Lord Kelvin had conjectured in 1887 that the division of space into equal-volume cells of least surface area is accomplished by a 14-sided polyhedron ("his minimal tetrakaidecahedron"). Physicists Weaire and Phelan exhibited a pair of polyhedra, one 12-sided and one 14-sided, with 0.3% less surface area. This solution is the basis for the repeating pattern in the building, which is "very flexible and thus efficient at absorbing seismic energy." Meanwhile, there is as yet no proof that Weaire and Phelan's solution is indeed minimal.

Nahin, Paul J., *Digital Dice: Computational Solutions to Practical Probability Problems*, Princeton University Press, 2008; xi + 263 pp, $27.95. ISBN 978-0-691-12698-2.

In the spirit of his earlier *Duelling Idiots and Other Probability Puzzlers* (2000), Nahin offers 21 very interesting (if not quite as practical as the title promises) probability problems and their solutions. Where possible, he gives an analytic solution, but the main thrust of this book is to display the power of the Monte Carlo method, as exhibited through simple Matlab code.

Gold, Bonnie, and Roger A. Simons (eds.), *Proof and Other Dilemmas: Mathematics and Philosophy*, MAA 2008; xxxii + 346 pp, $53.95 ($43.50 to members). ISBN 978-0-88385-567-6.

This book aims to increase interest in the philosophy of mathematics by offering new essays that concentrate on developments over the past 30 years. Sections include views on mathematical objects, mathematical knowledge, proof, social constructivism, and the applicability of mathematics. This is a valuable sourcebook for an undergraduate seminar; it is up to date, requires little specific mathematical knowledge, has numerous references for each essay, and—most importantly—has been written and edited carefully to avoid the obscurantism that plagues philosophy. (At the end of the last essay, Alan Hájek gives a novel way to score multiple-choice questions, based on the respondent assigning probabilities of correctness to each answer.)

Hidetoshi, Fukagawa, and Tony Rothman, *Sacred Mathematics: Japanese Temple Geometry*, Princeton University Press, 2008; xxviii + 348 pp, $35. ISBN 978-0-691-12745-3.

This book is the most thorough (and beautiful) account of Japanese temple geometry (*sangaku*) available. It traces the history of Japanese mathematics and culture that culminated in the placement in temples during the Edo period of artistic tablets displaying geometry problems. The book reproduces a number of the tablets in color plates, collects more than 100 problems (giving provenience and solution), and contains a translation of part of a travel diary of mathematician Yamaguchi Kanzan in the early 19th century (who copied down problems from temples he visited). The authors explain what is different about *sangaku* problems: They are "frequently more intricate than the usual exercises American students encounter. Instead of running four or five lines, proofs may run four or five pages—if not ten. What is more, it is necessary to bring to bear *everything* you've learned from your geometry course ... virtually every theorem ...."

Kendig, Keith, *Sink or Float? Thought Problems in Math and Physics*, MAA 2008; xiii + 375 pp, $59.95 ($47.95 to MAA members). ISBN 978-0-88385-339-9.

This is a collection of multiple-choice problems that are concrete and rooted in physical reality, and most can be solved by common sense (plus sometimes some calculation). Topics are geometry, numbers, astronomy, Archimedes' principle, probability, classical mechanics, electricity and magnetism, heat and wave phenomena, the leaking tank, and linear algebra. Illustrations abound, and solutions occupy almost half the book. My favorite problem was the one about the stereo camera (p. 19).

Dahlke, Richard M., *How to Succeed in College Mathematics: A Guide for the College Mathematics Student*, BergWay Publ., 2008; xv + 622 pp, $27.95 (P). ISBN 978-0-615-16803-6.

At first glance, I was skeptical of the length of this book (and the duplication in the subtitle); but it is chock full of clear exposition of excellent advice, with careful sectioning and bolding of important points, and it would be hard to point to anything to omit. Much of the advice applies to college in general, but much is specific to mathematics, including how courses are sequenced, how to get credit by examination, and how to obtain assistance. Author Dahlke explains the benefits of learning mathematics and what it means to think mathematically, as well as how to read, write, discuss, listen to, and work problems in mathematics. Rationales accompany the advice and thus render it more credible; Dahlke makes it clear how and why college differs from high school. I have two sons approaching college; I'll be giving them this book.

Demaine, Erik D., and Joseph O'Rourke, *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*, Cambridge University Press, 2007; xiii + 472 pp, $95, $49.99 (P). ISBN 978-0-521-85757-4, 978-0-521-71522-5. Woehr, Jack, A conversation with Erik Demaine, *Dr. Dobb's Journal* 33 (9) (September 2008) 16–18.

Erik Demaine's Hedrick Lectures at August's MathFest gave a delightful and masterful summary overview of how mathematics has been applied to novel areas of geometry, such as origami, linkages, hinged dissections, and transformer figures. His beautiful book, with its full-color illustrations, gives some of the details, including proofs and open problems.

NEWS AND LETTERS

Carl B. Allendoerfer Awards—2008

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.

Eugene Boman, Richard Brazier, and Derek Seiple, Mom! There's an Astroid in My Closet! MATHEMATICS MAGAZINE **80** (2007) 104–111.

Citation

If you have a cramped living space, your closet may have a bifold door; such a door requires less floor space. Few people ask, "How much less?" One morning a student dressing for school thought of it, asked two of his college professors, and the result is this elegant paper that connects modeling, calculus, the Trammel of Archimedes, and envelopes of families of ellipses. The authors investigate and solve the original problem, showing that the outer envelope of the area traced out by the closing door is bounded by two curves. One of the curves is the arc of a circle, but the other is the astroid, a curve with parametric representation $(x(\theta), y(\theta)) = (r\cos^3\theta, r\sin^3\theta)$. The authors solve extensions of the problem to doors with n folds with panels of varying lengths, and show that the asteroid keeps turning up. It is a curve that lives in many closets.

Biographical Notes

Eugene Boman received his BA from Reed College in 1984, his MA and PhD from the University of Connecticut in 1986 and 1993 respectively. He has been at The Pennsylvania State University (first at the DuBois campus and currently at the Harrisburg campus) since 1996.

Richard Brazier received a BSc Honors degree in Mathematics from Bath University England in 1991, his MS in 1994 and his PhD in 1997 in Applied Mathematics from University of Arizona. He is currently an Associate Professor of Mathematics and Geology at The Pennsylvania State University, DuBois campus and Chair of the DuBois campus Earth Science Program.

Derek Seiple received his BA from The Pennsylvania State University in 2007. He is currently working towards his PhD at the University of Arizona.

Response from Boman, Brazier, and Seiple

Naturally, the true reward was finding an interesting problem, working on it, and eventually solving it. However, we are truly surprised, amazed, humbled, and grateful that so many others seem to have found the problem and our solution of it interesting as well.

Chris Christensen, Polish Mathematicians Finding Patterns in Enigma Messages, MATHEMATICS MAGAZINE **80** (2007) 247–273.

Citation

This article is a well-crafted and carefully written historical paper about Enigma coding and how a little elementary mathematics can go a long way. The Enigma device was a cipher machine used by the Germans before and during World War II. The article describes the history, design, and workings of the device, as well as an account of the breaking of Enigma—a spectacular application of elementary mathematics. This occurred in the 1930s, when a team of Polish codebreakers, chiefly Marian Rejewski, found patterns in German Enigma messages. Their work was made possible by the fact that because of radio noise, each message setting was double-enciphered. This error, together with five elementary theorems about the cycle structure of permutations, allowed the Polish cryptographers to reduce the number of Enigma keys to be tested from seven quadrillion to about one hundred thousand. The result was a simplified model of Enigma which, with hard work and luck, allowed them to determine the internal settings of the machine.

In a book review in the *Notices of the AMS*, Jim Reeds has stated that "if ever there was a real-world story problem handed to mathematics teachers on a silver platter, this would be it." This article puts a shine on that platter.

Biographical Note

Chris Christensen is a professor of mathematics at Northern Kentucky University. His mathematical genealogy is a long line of algebraic geometers; he is a student of Professor S. S. Abhyankar. He became hooked on cryptology when, after running out of (good) Tom Clancy books, he read the novel Enigma by Robert Harris—a story set among the World War II British codebreakers at Bletchley Park.

Response from Christensen

I am deeply honored to receive an Allendoerfer Award. I enjoy exploring the impact of mathematicians and mathematics on cryptology, and I appreciate the MAA's encouragement of expository and historical writing by their publishing such papers in MATHEMATICS MAGAZINE. I would like to thank Professor Abhyankar, who has encouraged my writing style; Northern Kentucky University's Department of Mathematics, which has encouraged my cryptological exploration; and my wife, Nancy, who has been willing to accompany me on (many) trips to Bletchley Park, the National Cryptological Museum, etc., and who has put up with my cryptological obsessions.

UNITED STATES POSTAL SERVICE

Statement of Ownership, Management, and Circulation (Requester Publications Only)

1. Publication Title: Mathematics Magazine
2. Publication Number: 0025-570X
3. Filing Date: 9/2/08
4. Issue Frequency: Bimonthly Except July & August
5. Number of Issues Published Annually: Five
6. Annual Subscription Price (if any):
7. Complete Mailing Address of Known Office of Publication (Not printer) (Street, city, county, state, and ZIP+4): 1529 18th ST, NW WASHINGTON D.C. 20036
Contact Person: H. WALDMAN
Telephone (Include area code): 202-387-5200
8. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer): MAA, 1529 18th ST, NW, WASHINGTON, D.C. 20036
9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank)
Publisher (Name and complete mailing address): Math Assn of America, 1529 18th ST, NW, WASHINGTON, D.C. 20036
Editor (Name and complete mailing address): Allen Schwenk, Western Mich U, Kalamazoo, MI 49008
Managing Editor (Name and complete mailing address): H. WALDMAN, 1529 18th ST, NW, WASHINGTON, D.C. 20036
10. Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)

Full Name	Complete Mailing Address
MAA	1529 18th ST, NW, WASH, DC 20036

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address

12. Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

PS Form 3526-R, September 2007 (Page 1 of 3 (Instructions Page 3)) PSN: 7530-09-000-8855 PRIVACY NOTICE: See our privacy policy on www.usps.com

13. Publication Title: Mathematics Magazine
14. Issue Date for Circulation Data Below: June, 2008

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)		10,000	10,000
b. Paid Circulation (By Mail and Outside the Mail)	(1) Mailed Outside-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	9,500	9,500
	(2) Mailed In-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	0	0
	(3) Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®	0	0
	(4) Paid Distribution by Other Classes of Mail Through the USPS (e.g. First-Class Mail®)	0	0
c. Total Paid Distribution (Sum of 15b (1), (2), (3), and (4))		9,500	9,500
d. Free or Nominal Rate Distribution (By Mail and Outside the Mail)	(1) Free or Nominal Rate Outside-County Copies included on PS Form 3541	500	500
	(2) Free or Nominal Rate In-County Copies Included on PS Form 3541	0	0
	(3) Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g. First-Class Mail)	0	0
	(4) Free or Nominal Rate Distribution Outside the Mail (Carriers or other means)	0	0
e. Total Free or Nominal Rate Distribution (Sum of 15d (1), (2), (3) and (4))		500	500
f. Total Distribution (Sum of 15c and 15e)		10,000	10,000
g. Copies not Distributed (See Instructions to Publishers #4 (page #3))		0	0
h. Total (Sum of 15f and g)		10,000	10,000
i. Percent Paid (15c divided by 15f times 100)		95%	95%

16. Publication of Statement of Ownership
☒ If the publication is a general publication, publication of this statement is required. Will be printed in the October 2008 issue of this publication.
☐ Publication not required

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: [signature]
Date: 9/2/08

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

PS Form 3526, September 2007 (Page 2 of 3)

CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, NW
Washington, DC 20036