LETTER FROM THE EDITOR

The issue begins with an article by Kady Hossner Boden and Michael Ward, in which they construct Cayley-Sudoku tables. Their work applies a theorem from 1939 by Reinhold Baer and another from 1967 by Jozsef Dénes. This article also inspired David Reimann's cover art, which is a representation of one of the Sudoku tables from the lead article. The cover art uses a method of representing permutations as closed curves that was developed by Karl Kattchee and Craig Kaplan; see the inside front cover for more information about the cover piece.

In the next article, Saran Ishika Maiti and Jyotirmoy Sarkar consider two unusual amusement park rides to introduce symmetric simple random walks on finite lattices. They answer questions about the likelihood to return to a starting position and the number of nodes visited, but do so using elementary ideas about symmetry, recursive relations, and induction.

In the *Triphos* system, the concept of number is based on three attributes. Because cancellation occurs when all three of these attributes have the same magnitude, there is no need for subtraction. Keely Grossnickle, Brian Hollenback, Jeana Johnson, and Zhihao Sun prove a number of results about the *Triphos* system and conclude by discussing some open questions that may be suitable for students to answer.

Applied mathematics often means the application of mathematics to solve a problem in another discipline. But, in the next article, graph theory is applied to answer a question about the roots of polynomials over finite fields. More specifically, Robert S. Coulter, Stefaan De Winter, Alex Kodess, and Felix Lazebnik prove that certain trinomials over a finite field have the same number of distinct roots. Their proof hinges on an isomorphism of related directed graphs.

In 1837, Pierre Wantzel's provided an algebraic proof that it is impossible to construct a cube whose volume is exactly twice the volume of a given cube using only a straightedge and a compass. Wantzel's proof involves the algebraic properties of constructible numbers and the irreducibility of a specific cubic polynomial. In the next article, Matt Lunsford extends the notion of Platonic solids being constructible over prime finite fields.

This issue is full of visuals. Mixed throughout this issue are proofs without words by Charles F. Marion; Sanja Stevanović and Dragan Stevanović; Mingjang Chen; John Molokach; Vincent Ferlini; Tom Edgar; and Rex H. Wu. Allison Henrich interviews Bob Bosch about his use of optimization to create artwork. While the interview includes images of his work, there are five other images sprinkled throughout the issue, including his piece *Embrace* and four emoji renditions of popular Halloween and other images.

The issue concludes with the Reviews, Problems, and announcement of the Allendoerfer Awards. One more thing, there is a correction to Problem 2067 from the April 2019 issue. Please see the Problems section in this issue for more details.

Michael A. Jones, Editor

ARTICLES

A New Class of Cayley-Sudoku Tables

KADY HOSSNER BODEN
St. Stephen's Academy
Beaverton, OR 97008
kboden@ststephensacademy.com

MICHAEL B. WARD
Western Oregon University
Monmouth, OR 97361
wardm@wou.edu

A Cayley-Sudoku table of a finite group G is a Cayley (i.e. operation) table for G subdivided into uniformly sized rectangular blocks, in such a way that each group element appears once in each block. For example, Table 1 is a Cayley-Sudoku table for $\mathbb{Z}_9 := \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ under addition modulo 9 and Table 2 is a Cayley-Sudoku table for S_3, the symmetric group on three symbols. Obviously inspired by the popularity of Sudoku puzzles, Cayley-Sudoku tables were introduced in [2], which gave three constructions. The second construction involved a curious condition. After several fruitless inquiries, the second author asked about that condition in a talk at the XXXth Ohio State-Denison Math Conference where, fortuitously, quasigroup theorists were in attendance. Numerous hands went up. First responder Professor Clifton E. Ealy, Jr. of Western Michigan University announced "You and your students have rediscovered a 1939 theorem of Reinhold Baer!"

In this note, we first review Constructions 1 and 2 of [2], then we briefly explain the "rediscovered" theorem of Baer [1], all of which requires only a familiarity with elementary group theory through cosets. Next we turn to our principal result, new instances of Construction 2 gleaned from Baer, which is more technical and requires some experience with permutation groups. These instances answer an open question in [2] about when the aforementioned curious condition is met. We conclude with a remark concerning Construction 1 and a theorem of József Dénes [3, 4].

	0	3	6	1	4	7	2	5	8
0	0	3	6	1	4	7	2	5	8
1	1	4	7	2	5	8	3	6	0
2	2	5	8	3	6	0	4	7	1
3	3	6	0	4	7	1	5	8	2
4	4	7	1	5	8	2	6	0	3
5	5	8	2	6	0	3	7	1	4
6	6	0	3	7	1	4	8	2	5
7	7	1	4	8	2	5	0	3	6
8	8	2	5	0	3	6	1	4	7

TABLE 1: $\mathbb{Z}_9$ Cayley-Sudoku table.

Math. Mag. **92** (2019) 243–251. doi:10.1080/0025570X.2019.1613949
MSC: Primary 05B15

	(1)	(13)	(132)	(12)	(123)	(23)
(1)	(1)	(13)	(132)	(12)	(123)	(23)
(12)	(12)	(123)	(23)	(1)	(13)	(132)
(13)	(13)	(1)	(12)	(132)	(23)	(123)
(132)	(132)	(23)	(123)	(13)	(1)	(12)
(23)	(23)	(132)	(13)	(123)	(12)	(1)
(123)	(123)	(12)	(1)	(23)	(132)	(13)

TABLE 2: S_3 Cayley-Sudoku table.

Constructing Cayley-Sudoku tables

Tables 2 and 1 illustrate Construction 1 of [2]. For Table 2, consider the subgroup $S = \langle(12)\rangle = \{(1), (12)\}$. Notice the rows in each block of the table are labeled with elements of the left cosets $(1)S = \{(1), (12)\}$, $(13)S = \{(13), (132)\}$, and $(23)S = \{(23), (123)\}$. (Permutations here are composed left to right.) The right cosets are $S(1) = \{(1), (12)\}$, $S(13) = \{(13), (123)\}$, and $S(23) = \{(23), (132)\}$ and the columns in each block are labeled with a complete set of right coset representatives of S, $R_1 = \{(1), (13), (132)\}$ and $R_2 = \{(12), (123), (23)\}$ (that is, one element from each right coset).

In the sequel, we usually economize by indicating the row and column labels for the blocks by listing the sets of labels rather than individual labels. Thus, the condensed layout of Table 2 has rows labeled $(1)S$, $(13)S$, and $(23)S$ and columns labeled R_1 and R_2.

The following theorem, which we call a "construction" for consistency with [2], simply says that such a layout always produces a Cayley-Sudoku table, as does its dual, obtained by switching right with left and rows with columns.

Construction 1 ([2], p. 133)**.** *Let G be a finite group and S be a subgroup of G with order k and index n. If $y_1S, y_2S, \ldots, y_nS$ are the n distinct left cosets of S in G, then arranging the Cayley table of G with rows labeled by the cosets $y_1S, y_2S, \ldots, y_nS$ and the columns labeled by sets $R_1, R_2, \ldots, R_k$ (as in Table 3, Left) yields a Cayley-Sudoku table of G with blocks of dimension $k \times n$ if and only if $R_1, R_2, \ldots, R_k$ partition G into complete sets of right coset representatives of S in G.*

Furthermore, if $Sg_1, Sg_2, \ldots, Sg_n$ are the distinct right cosets of S in G, then arranging the Cayley table of G with columns labeled by the cosets $Sg_1, Sg_2, \ldots, Sg_n$ and the rows labeled by sets $L_1, L_2, \ldots, L_k$ (as in Table 3, Right) yields a Cayley-Sudoku table of G with blocks of dimension $n \times k$ if and only if $L_1, L_2, \ldots, L_k$ partition G into complete sets of left coset representatives of S in G.

We refer to the first part as Construction 1L because it uses left cosets and the second one as 1R. Construction 1R is illustrated by Table 1 using the subgroup $S = \langle 3\rangle = \{0, 3, 6\}$ of $\mathbb{Z}_9$. For any subgroup of any finite group, one can always partition the group into complete sets of left or right coset representatives. Therefore, every group has Cayley-Sudoku tables corresponding to each of its subgroups.

We now turn to the next construction, for which we review a definition. If S is a subgroup of the group G and $g \in G$, S^g denotes the subgroup $g^{-1}Sg := \{g^{-1}sg : s \in S\}$, which is called a *conjugate* of S.

Construction 2 ([2], p. 134)**.** *Let S be a subgroup of G with order k and index n. Also suppose $y_1S, y_2S, \ldots, y_nS$ are the distinct left cosets of S in G. Arranging the Cayley*

table of G with columns labeled by the cosets y_1S, y_2S, ..., y_nS and the rows labeled by sets $L_1, L_2, \ldots, L_k$ (as in Table 4, Left) yields a Cayley-Sudoku table of G with blocks of dimension $n \times k$ if and only if $L_1, L_2, \ldots, L_k$ are complete sets of left coset representatives of S^g for all $g \in G$.

Furthermore, suppose Sg_1, Sg_2, ..., Sg_n are the distinct right cosets of S in G. Arranging the Cayley table of G with rows labeled by the cosets Sg_1, Sg_2, ..., Sg_n and columns labeled by the sets R_1, R_2, ... R_k (as in Table 4, Right) yields a Cayley-Sudoku table of G if and only if R_1, R_2, ... R_k partition G into complete sets of right coset representatives of S^g for all $g \in G$.

	R_1	R_2	...	R_k
y_1S				
y_2S				
$\vdots$				
y_nS				

	Sg_1	Sg_2	...	Sg_n
L_1				
L_2				
$\vdots$				
L_k				

TABLE 3: (Left) Construction 1L using left cosets and right coset representatives. (Right) Construction 1R using right cosets and left coset representatives.

	y_1S	y_2S	...	y_nS
L_1				
L_2				
$\vdots$				
L_k				

	R_0	R_1	$\cdots$	R_k
Sg_1				
Sg_2				
$\vdots$				
Sg_m				

TABLE 4: (Left) Construction 2L using left cosets and left coset representatives. (Right) Construction 2R using right cosets and right coset representatives

We refer to the two parts as Construction 2L and 2R, respectively, referring again to the use of left or right cosets.

To illustrate Construction 2L, let $G = S_3$ and $S = \langle (1, 2) \rangle$. The left cosets are $(1)S = \{(1), (12)\}$, $(13)S = \{(13), (132)\}$, and $(23)S = \{(23), (123)\}$. The conjugates of S in G are $\langle (12) \rangle$, $\langle (13) \rangle$, and $\langle (23) \rangle$. It is easy to check that $L_1 = \{(1), (123), (132)\}$ and $L_2 = \{(12), (13), (23)\}$ partition G into complete sets of left coset representatives for each of the conjugates of S in G. Thus, Table 5 is an instance of Construction 2L.

Construction 2L looks very similar to Construction 1R, but 1R required *right* cosets and left coset representatives. The price we pay for using *left* cosets along with left coset representatives in 2L is that the coset representatives must be complete sets of left coset representatives not just for the subgroup S but for all the conjugates of S at once. This can be a high price. For example, it is impossible to find such representatives for the subgroup $\langle (12)(34) \rangle$ in the symmetric group S_4. The obvious question "When *can* we get such coset representatives?" led to the enthusiastic reference to Baer's theorem mentioned in the introduction.

	(1)	(12)	(13)	(132)	(23)	(123)
(1)	(1)	(12)	(13)	(132)	(23)	(123)
(123)	(123)	(23)	(12)	(1)	(13)	(132)
(132)	(132)	(13)	(23)	(123)	(12)	(1)
(12)	(12)	(1)	(123)	(23)	(132)	(13)
(13)	(13)	(132)	(1)	(12)	(123)	(23)
(23)	(23)	(123)	(132)	(13)	(1)	(12)

TABLE 5: Another S_3 Cayley-Sudoku Table

Baer's theorem and Construction 2

A set Q with a binary operation $\cdot$ is a *quasigroup* provided its Cayley table is a bordered Latin square on the elements of Q or, equivalently, it has cancellation, which is to say, for every $x, y, a \in Q$, if $a \cdot x = a \cdot y$, then $x = y$ (row a in the Cayley table contains each element only once) and $x \cdot a = y \cdot a$ implies $x = y$ (column a in the Cayley table contains each element only once). Removing the borders of Table 6, for example, leaves a Latin square, so it gives a quasigroup. In a quasigroup, there need not be inverses, an identity, or even associativity.

With S and G as in Construction 2R, fix a complete set of right coset representatives $\{r_1, r_2, \ldots, r_m\}$ of S in G. (Think of this as one of the R_i in Construction 2.) Define a coset multiplication by $Sr_i \cdot Sr_j := Sr_ir_j$. Because the coset representatives are fixed, this gives a binary operation on $\mathcal{R}$, the set of right cosets of S in G. Normally, coset multiplication like that is well-defined only when S is a normal subgroup of G. However, by fixing the coset representatives in advance, all is well. We can now state the 1939 theorem mentioned in the introduction ([**1**], Theorem 2.3, where "division system" is used in place of "quasigroup").

Theorem (Baer's Theorem)**.** *$\mathcal{R}$ under $\cdot$ as defined above is a quasigroup if and only if $\{r_1, r_2, \ldots, r_m\}$ is a complete set of right coset representatives of S^g for every $g \in G$.*

That last bit is the same condition as in Construction 2R. Therefore, we have the following theorem.

Theorem (Construction 2R à la Baer)**.** *With notation as in Construction 2R, the following are equivalent.*

(a) The arrangement of the Cayley table in Table 4 (Right) gives a Cayley-Sudoku table.

(b) The sets R_1 through R_k partition G and each is a complete set of right coset representatives of S^g for every $g \in G$.

(c) The sets R_1 through R_k partition G and each gives rise to a quasigroup on the right cosets of S as described above.

In other words, Construction 2R is Baer's theorem in disguise. We chose Construction 2R in order to conform with Baer's use of right cosets. One can also prove the left-handed version corresponding to Construction 2L.

New instances of Construction 2

Other than the trivial case where S is a normal subgroup (and Construction 2 reduces to Construction 1), the following two propositions gave the only general setting known

to the authors of [2] wherein Construction 2 could be applied. The authors asked for other such settings ([2], p. 135).

Proposition 1. *Assume S is a subgroup of a finite group G. Suppose R is a complete set of right [left] coset representatives of S^g for all $g \in G$. Then the sets $sR := \{sr : r \in R\}$ $[Rs := \{rs : r \in R\}]$, $s \in S$ partition G into complete sets of right [left] coset representatives of S^g for all $g \in G$.*

In other words, in applying Construction 2, it is sufficient to find one set of coset representatives of the desired sort. The next proposition gives one setting where such a set exists.

Proposition 2. *Suppose S is a subgroup of the finite group G and there is a subgroup C such that $G = CS$ and $C \cap S$ is the trivial subgroup (i.e. C is a complement for S), then, for all $g \in G$, C is a complement for S^g in G and C is a complete set of left and right coset representatives of S^g in G.*

Table 5 is an instance of Propositions 2 and 1 since L_1 is a complement for S in G and $L_2 = L_1(12)$.

While Baer's theorem gives another way to think about Construction 2, it does not directly give the new instances of the construction called for by the authors of [2], "new" meaning not accounted for by Proposition 2. Nevertheless, reading Baer revealed a new class of examples arising from quasigroups, which we will now explain, beginning with a quick overview of some basics about permutation groups and quasigroups.

Suppose G is a group of permutations of a set A. For each $a \in A$ and $g \in G$, a^g denotes the image of a under g and $G_a := \{g \in G : a^g = a\}$ denotes the *stabilizer* of a in G. (We beg the reader's pardon for writing a^g in place of the more familiar function notation $g(a)$. It is common practice among group theorists, and it is the notation used in GAP [6] calculations.) G is *transitive* when for every $a, b \in A$ there is a $g \in G$ such that $a^g = b$. G is *regular* provided G is transitive and G_a is the trivial subgroup for any $a \in A$.

The following is a standard result. Its proof may be found in any treatment of permutation groups (or may be taken on as an exercise by the reader).

Proposition 3. *Suppose G is a transitive group of permutations on a set A and $a \in A$.*
(a) For each $g \in G$, $(G_a)^g := g^{-1}G_a g = G_{a^g}$.
(b) If T is a regular subgroup of G, then $|T| = |A|$.
(c) If T is a complement of G_a in G, then T is regular.

Suppose Q with operation $\cdot$ is a finite quasigroup. For each $\ell \in Q$ define $\rho_\ell, \lambda_\ell : Q \to Q$ by $q^{\rho_\ell} = q \cdot \ell$ and $q^{\lambda_\ell} = \ell \cdot q$. Using the definition of quasigroup, one can easily prove the next result.

Proposition 4. *Suppose Q is a finite quasigroup. For each $\ell \in Q$, ρ_ℓ and λ_ℓ are permutations of Q. For every $a, b \in Q$ there exist $\ell, l \in Q$ such that $a^{\rho_\ell} = b$ and $a^{\lambda_l} = b$.*

When Q is a group, $\{\rho_\ell : \ell \in Q\}$ and $\{\lambda_\ell : \ell \in Q\}$ are permutation groups isomorphic to Q. That is Cayley's representation theorem. In general, however, those sets are not groups. Instead, quasigroup theorists consider the groups of permutations generated by those sets, $RMult(Q) := \langle \rho_\ell : \ell \in Q \rangle$ and $LMult(Q) := \langle \lambda_\ell : \ell \in Q \rangle$. By Proposition 4, each of these groups is transitive.

We can now describe instances of Construction 2 arising from quasigroups.

Proposition 5. *Suppose Q is a quasigroup and $c \in Q$. Let $G = RMult(Q)$, then $R := \{\rho_\ell : \ell \in Q\}$ is a complete set of right coset representatives of $(G_c)^g$ for every $g \in G$. Similarly, if $G = LMult(Q)$, then $L := \{\lambda_\ell : \ell \in Q\}$ is a complete set of left coset representatives of $(G_c)^g$ for every $g \in G$.*

Proof. Let $G = RMult(Q)$ and $g, h \in G$. By Proposition 4, there exists $\ell \in Q$ such that $(c^g)^{\rho_\ell} = c^{gh}$. Therefore, $\rho_\ell h^{-1} \in G_{c^g}$ which equals $(G_c)^g$ by Proposition 3. Thus, $(G_c)^g \rho_\ell = (G_c)^g h$ and R contains a representative of each coset of $(G_c)^g$. Moreover, for $\ell, m \in Q$, $(G_c)^g \rho_\ell = (G_c)^g \rho_m$ if and only if $(c^g)^{\rho_\ell} = (c^g)^{\rho_m}$, that is, $(c^g) \cdot \ell = (c^g) \cdot m$. Therefore, $\ell = m$ by cancellation in Q. Thus, R is a complete set of right coset representatives of $(G_c)^g$ for every $g \in G$. The proof for $G = LMult(Q)$ is similar. ∎

Proposition 5 and Proposition 1 imply that for any quasigroup Q, Construction 2 applies to $G = LMult(Q)$ and to $G = RMult(Q)$ using the subgroup G_c for any $c \in Q$ and coset representatives L and R, respectively. Some of these lead to new examples of Construction 2, where G_c does not have a complement, as we now illustrate. First, for the convenience of the reader, we prove a well-known proposition.

Proposition 6. *Suppose n is a positive integer and $n \equiv 2 (mod\ 4)$, then the alternating group A_n does not contain a regular subgroup.*

Proof. With n as in the hypotheses, assume T is a regular subgroup of A_n. By Proposition 3, $|T| = n$. Thus, T contains an element t of order 2 by Cauchy's theorem. Since T is regular, t has no fixed points. Therefore, it is the composition of $n/2$ 2-cycles, a contradiction since $n/2$ is odd. ∎

Example 1. Table 6 defines a quasigroup Q_6 since the table is visibly a bordered Latin square. We calculate $\lambda_1 = (1)$, $\lambda_2 = (123)(456)$, $\lambda_3 = (132)(465)$, $\lambda_4 = (14)(2536)$, $\lambda_5 = (15)(2634)$, and $\lambda_6 = (16)(2435)$, all even permutations.

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	3	1	5	6	4
3	3	1	2	6	4	5
4	4	5	6	1	3	2
5	5	6	4	2	1	3
6	6	4	5	3	2	1

TABLE 6: Quasigroup Q_6

Therefore, $G = LMult(Q_6)$ is a subgroup of the alternating group A_6. Suppose G_1 has a complement C. Then C is a regular subgroup by Proposition 3, contradicting Proposition 6. Thus, applying Construction 2L (and Proposition 1) to G using the subgroup G_1 and the left coset representatives $L = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6\}$ gives a new instance of the construction.

By the way, GAP [**6**] tells us G has order 36 and

$$G_1 = \{(1), (456), (465), (23)(56), (23)(45), (23)(46)\}.$$

The corresponding Cayley-Sudoku Table from Construction 2L has blocks of dimension 6×6.

Also, $\rho_1 = (1)$, $\rho_2 = (123)(456)$, $\rho_3 = (132)(465)$, $\rho_4 = (14)(25)(36)$, $\rho_5 = (15)(26)(34)$, and $\rho_6 = (16)(24)(35)$ are not all even permutations. Using GAP again, $G = RMult(Q_6)$ has order 18 and $G_1 = \{(1), (456), (465)\}$. The corresponding Cayley-Sudoku table from Construction 2R has blocks of dimension 6×3. However, G_1 does have a complement, namely $\langle \rho_2 \rho_4 \rangle = \langle (153426) \rangle$, so, unfortunately, this does not give a new instance of Construction 2.

Example 2. Example 1 readily generalizes to any quasigroup Q_n of order $n > 2$ where $n \equiv 2 (\mathrm{mod}\ 4)$ and the left (or right) translations are even permutations. Table 7 illustrates just such a generalization. To verify it is a quasigroup, consider the four subtables formed by the dashed lines. The upper left and lower right subtables contain the numbers 1 through $\frac{n}{2}$ while the other two contain the numbers $\frac{n}{2} + 1$ through n. Moreover, each successive row in the lower right subtable is the previous row shifted (with wrapping) one place to the right. In the remaining subtables, rows are shifted one place to the left. Thus we see the table is a bordered Latin square.

One calculates λ_i to be the permutation $\left((1, 2, \ldots, \frac{n}{2})(\frac{n}{2} + 1, \frac{n}{2} + 2, \ldots, n)\right)^{i-1}$ for $1 \le i \le \frac{n}{2}$, and $(1, i)(2, i + 1, 3, i + 2, \ldots, \frac{n}{2} - 1, i + \frac{n}{2} - 2, \frac{n}{2}, i + \frac{n}{2} - 1)$ for $\frac{n}{2} + 1 \le i \le n$ where addition is done modulo $n/2$. It is not hard to see each λ_i is an even permutation under the hypotheses on n.

Other examples are possible. As long as the left or right translations lie in any permutation group not having a regular subgroup, no complement exists. It is interesting to know such groups exist of every order except 1, a prime, or a prime squared ([7], Theorem 4).

Dénes's theorem and Construction 1

It turns out that Construction 1 is also a rediscovery. This time of a theorem of Dénes [3, 4].

An $(m, 1)$-*complete Latin rectangle* is a rectangle that can be completed to a Latin square and contains m different symbols each occurring exactly once. Since every Cayley table is a (bordered) Latin square, the blocks in our Cayley-Sudoku tables are $(m, 1)$-complete Latin rectangles where m is the order of the group. Dénes stated the following theorem.

Theorem ([4], Theorem 1.5.5)**.** *If L is the Latin square representing the multiplication [Cayley] table of a group G of order m, where m is a composite number, then L can be split [partitioned] into a set of m* $(m, 1)$*-complete non-trivial [i.e. not consisting of a single row or column] Latin rectangles.*

The "split" table is clearly a Cayley-Sudoku table of G! Moreover, in proving the theorem, Dénes takes a subgroup of G and arranges the Cayley table exactly as specified in Construction 1L–with one possible flaw. The coset representatives he designates as column labels might not be right coset representatives as required. Use of right versus left cosets is ambiguous in the proof and the examples in [3, 4] use normal subgroups where the distinction is irrelevant. Nevertheless, Dénes gets credit for constructing the first Cayley-Sudoku table. The authors of [2] might take solace in having repopularized the notion, since Theorem 1.5.5 is one of the "topics no longer of current interest" omitted in the second edition [5] of the classic text [4] in which it appeared.

(By the way, in his proof, Dénes uses a proper non-trivial subgroup, which exists since m is composite. The use of a proper non-trivial subgroup is only to ensure the non-triviality of the Latin rectangles or blocks, as we call them. Since blocks consisting of a single row or column are allowed in [2], Construction 1 does not assume that the order of the group is composite.)

	1	2	3	$\cdots$	$\frac{n}{2}-1$	$\frac{n}{2}$	$\frac{n}{2}+1$	$\frac{n}{2}+2$	$\frac{n}{2}+3$	$\cdots$	$n-1$	n
1	1	2	3	$\cdots$	$\frac{n}{2}-1$	$\frac{n}{2}$	$\frac{n}{2}+1$	$\frac{n}{2}+2$	$\frac{n}{2}+3$	$\cdots$	$n-1$	n
2	2	3	4	$\cdots$	$\frac{n}{2}$	1	$\frac{n}{2}+2$	$\frac{n}{2}+3$	$\frac{n}{2}+4$	$\cdots$	n	$\frac{n}{2}+1$
3	3	4	5	$\cdots$	1	2	$\frac{n}{2}+3$	$\frac{n}{2}+4$	$\frac{n}{2}+5$	$\cdots$	$\frac{n}{2}+1$	$\frac{n}{2}+2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$\frac{n}{2}$	$\frac{n}{2}$	1	2	$\cdots$	$\frac{n}{2}-2$	$\frac{n}{2}-1$	n	$\frac{n}{2}+1$	$\frac{n}{2}+2$	$\cdots$	$n-2$	$n-1$
$\frac{n}{2}+1$	$\frac{n}{2}+1$	$\frac{n}{2}+2$	$\frac{n}{2}+3$	$\cdots$	$n-1$	n	1	3	4	$\cdots$	$\frac{n}{2}$	2
$\frac{n}{2}+2$	$\frac{n}{2}+2$	$\frac{n}{2}+3$	$\frac{n}{2}+4$	$\cdots$	n	$\frac{n}{2}+1$	2	1	3	$\cdots$	$\frac{n}{2}-1$	$\frac{n}{2}$
$\frac{n}{2}+3$	$\frac{n}{2}+3$	$\frac{n}{2}+4$	$\frac{n}{2}+5$	$\cdots$	$\frac{n}{2}+1$	$\frac{n}{2}+2$	$\frac{n}{2}$	2	1	$\cdots$	$\frac{n}{2}-2$	$\frac{n}{2}-1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	n	$\frac{n}{2}+1$	$\frac{n}{2}+2$	$\cdots$	$n-2$	$n-1$	3	4	5	$\cdots$	2	1

TABLE 7: Quasigroup Q_n.

Acknowledgment This paper is an outgrowth of the first author's honors thesis at Western Oregon University, supervised by the second author.

REFERENCES

[1] Baer, R. (1939). Nets and groups. *Trans. Amer. Math. Soc.* 46: 110–141.

[2] Carmichael, J., Schloeman, K.,Ward, M. B. (2010). Cosets and Cayley-Sudoku tables. *Math. Mag.* 83(2): 130–139.

[3] Dénes, J. (1967). Algebraic and combinatorial characterizations of Latin squares I. *Math. Slovaca* 17: 249–265.

[4] Dénes, J., Keedwell, A. D. (1974). *Latin Squares and Their Applications*. New York: Academic Press.

[5] Keedwell, A. D., Dénes, J. (2015). *Latin Squares and Their Applications*. 2nd ed. Amsterdam: North-Holland.

[6] The GAP Group (2013). GAP – Groups, Algorithms, and Programming, Version 4.6.2. gap-system.org.

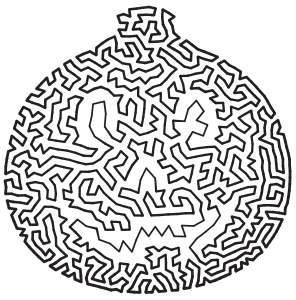
[7] Xu, M.-Y., (2008). A note on permutation groups and their regular subgroups. *J. Aust. Math. Soc.* 85: 283–287.

Summary. A Cayley-Sudoku table of a finite group G is a Cayley table for G subdivided into uniformly sized rectangular blocks, in such a way that each group element appears once in each block. Cayley-Sudoku tables and three ways to construct them were introduced in 2010 by J. Carmichael, K. Schloeman, and M. B. Ward. We give new instances of Cayley-Sudoku tables gleaned from a theorem of Reinhold Baer, which answer an open question of Carmichael, Schloeman, and Ward. We also briefly explain theorems of Baer and Jozsef Dénes and their connections to the constructions.

KADY HOSSNER BODEN (MR Author ID: 1311907) earned a B.S. with honors from Western Oregon University in 2011 and an M.A.T from Multnomah University in 2013. She is the K-12 Mathematics Coordinator at St. Stephens Academy where she enjoys teaching middle and high school students. In her spare time she is an avid birder and mystery lover.

MICHAEL B. WARD (MR Author ID: 207753) received a Ph.D. from the University of Utah in 1979. He spent the first half of his career at Bucknell University and the second half at Western Oregon University where he is Emeritus Professor of Mathematics. In retirement, he enjoys community service; spending extra time with his wife, children, and grandson; and not having papers to grade.

Embrace, Robert Bosch; 2009. Awarded First Prize at the AMS Mathematical Art Exhibition at the 2010 Joint Mathematics Meetings in San Francisco. One quarter inch thick and eight inches in diameter. The "inside" piece is stainless steel, and the "outside" piece is brass. The "gap" is an optimal solution to a 726-point instance of the TSP in which the points have 6-fold rotational symmetry and the tour was forced to have 3-fold rotational symmetry.
See interview on page 305.



Opt-emoji Jack-o'-Lantern, Robert Bosch; digital print, 2017. An optimal TSP tour of 1024 points that were arranged to resemble an emoji rendition of a jack-o'-lantern. The optimal tour was obtained with the Concorde TSP Solver.
See interview on page 305.

Symmetric Walks on Paths and Cycles

SARAN ISHIKA MAITI
Visva-Bharati University
Santiniketan, West Bengal, India 731235
saran.ishika@gmail.com

JYOTIRMOY SARKAR
Indiana University-Purdue University Indianapolis
Indianapolis, IN 46202
jsarkar@iupui.edu

The following rides from an unusual theme park motivate this paper.

The carnival elevator problem. On Carnival Land's Haunted High-rise, floors are numbered $G, 1, 2, \ldots, 10$. You enter an elevator on Floor G and then travel from one floor to another to experience a unique scary event on each floor. The elevator is a bit weird—once it stops on a floor, it forgets which floor it came from. After the riders experience the scary event, the elevator goes to an adjacent floor: From Floor G the elevator goes to Floor 1; from Floors $2, \ldots, 9$, it is equally likely to go up or down one floor; from Floor 10 it goes to Floor 9. Each floor-to-floor transition takes a negligible amount of time. When the elevator returns to Floor G all visitors must exit. If the elevator spends 2 minutes on each floor, how long will your haunted high-rise tour be? How many floors will you have visited when your tour ends? How many tours will it take on average until you visit the topmost Floor 10?

The carnival rotating multiplex problem. On Carnival Land's Rotating Multiplex, six theaters are numbered $0, 1, 2, \ldots, 5$. Each theater repeats a unique show every 5 minutes. You enter in Theater 0, the doors shut, you sit in your chair and enjoy the ongoing performance for 5 minutes before the theaters rotate clockwise or counter-clockwise at random to let you experience the performance in the adjacent theater. Doors open and you are allowed to exit only when you are back at Theater 0. How long will it be until you return to Theater 0? How many distinct shows will you have seen by then?

If you decide to sit in your chair until you have seen performances in all six theaters at least once, then how long will it take you to see all six shows on average? Which show will you see last (among these six shows)? At the beginning of the show in this last theater you will receive a complimentary hat unique to that Theater. Thereafter, you will exit the multiplex as soon as you face Theater 0. How much longer will you stay in your chair, and how many distinct shows will you see after receiving your hat until you exit?

Toward the end of this paper, we reveal the solutions to these problems. First we develop the techniques to study a symmetric simple random walk (henceforth simply called a walk) on a linear or a cyclic network. A random walk on any graph is a sequence of nodes that are visited one by one, starting from a given node, and at each step moving to one of the neighbors of the current node chosen according to some probability distribution. In this paper, we study simple random walks on finite sections of the integer lattice (with various end conditions) and on the vertices of a polygon,

Math. Mag. **92** (2019) 252–268. doi:10.1080/0025570X.2019.1611166
MSC: Primary 60J10; Secondary 05C81
Supplemental data for this article can be accessed on the publisher's website.

where each move takes the walk to one of the neighbors with equal probability. We are interested in studying the waiting time (or the number of steps needed) until some predefined event occurs and related questions.

Our goal is to present elementary solutions to the waiting time and related problems. While sometimes we refer to more advanced solutions available elsewhere, in this paper we utilize only elementary tools such as symmetry, recursive relations and mathematical induction. We assume the reader is familiar with a few probability concepts and results—including the law of total expectation (Adam's law) and the law of total variance (Eve's law). An undergraduate course in probability will suffice. We also assume the reader can prove some combinatorial identities. These laws and identities appear as a supplement on the publisher's website.

Why study random walks? The standard introductory probability course considers one random experiment at a time. You toss a coin, roll a die, deal cards from a deck, draw balls from an urn, etc.; and then you get a reward or a penalty depending on the outcome of the experiment. Analysis of these problems involves a finite sample space. Waiting time problems (which this paper treats) are conceptually quite different, because there is an infinite (or even an uncountable) sample space. Studying random walks is a good way to build intuition about waiting times.

Random walks have many applications: Physicists use random walks to study Brownian motion of particles in fluids and statistical mechanics. Computer scientists use random walks to model epidemic diffusion of information, or to generate random samples from large complex networks and to compute aggregate functions on complex sets. Often, the analyses boil down to studying the gambler's ruin problem, which is essential for our problems, too.

Gambler's ruin problem

The gambler's ruin problem may be viewed as a random walk on a graph with nodes $\{0, 1, \ldots, m\}$, and edges $(i-1, i)$ for $i = 1, 2, \ldots, m$. In particular, nodes 0 and m are absorbing (once the walk reaches either of them, it cannot escape from it); and each of the other nodes leads to its two neighbors with probability $1/2$ each. See Network A of Figure 1.

Theorem 1 (Gambler's ruin problem (GRP))**.** *A gambler enters a casino with x dollars to play a game repeatedly until either her fortune reaches m dollars (where $m > x$ is a fixed number) or she goes broke. For each play she either wins one dollar or loses one dollar, each with probability $1/2$. The following results hold:*

GRP(1). The probability that the gambler wins is x/m;
GRP(2). The number of plays until the game ends has expectation $x(m-x)$; and
GRP(3). The game duration has variance $x(m-x)\,[x^2 + (m-x)^2 - 2]/3$.

For a proof of the three results in Theorem 1, we refer the reader to Ross [**8**], Feller [**4**], and Good [**5**], respectively. There are also clever proofs using martingales in Karlin and Taylor [**6**], and using electrical theory in Chandra et al. [**1**] and Doyle and Snell [**3**]. However, to set the tone of this paper, we provide elementary proofs using only recursive relations and mathematical induction.

Proof. GRP(1). Let P_x denote the probability that starting from node x the walk reaches node m before it reaches node 0. Then $P_0 = 0$, $P_m = 1$, and for all $1 \le x \le m-1$, by conditioning on the first outcome, $P_x = (P_{x-1} + P_{x+1})/2$; or equivalently, $P_{x+1} - P_x = P_x - P_{x-1}$. Thus, the differences $P_1 - P_0$, $P_2 - P_1, \ldots, P_m - P_{m-1}$ are all equal. Adding all these m differences and substituting $P_m = 1$, $P_0 = 0$, we get

$m\,P_1 = 1$; or $P_1 = 1/m$. Finally, adding only the first x differences, we get $P_x = (P_1 - P_0) + (P_2 - P_1) + \cdots + (P_x - P_{x-1}) = x\,P_1 = x/m$.

GRP(2). Let T_x denote the waiting time (or the number of steps) the walk takes starting from node x to reach either node m or node 0, and let $E_x = E[T_x]$ denote the expected waiting time. Then $E_0 = 0 = E_m$, and for all $1 \le x \le m-1$, we have $E_x = 1 + (E_{x-1} + E_{x+1})/2$; or equivalently, $E_{x+1} - E_x = E_x - E_{x-1} - 2$. Thus, the successive differences $E_1 - E_0, E_2 - E_1, \ldots, E_m - E_{m-1}$ are in arithmetic progression with common difference -2. Adding all these m differences, we have

$$(E_1 - E_0) + (E_2 - E_1) + (E_3 - E_2) + \cdots + (E_m - E_{m-1}) = E_m - E_0.$$

Substituting $E_m = 0 = E_0$, we get

$$E_1 + (E_1 - 2) + (E_1 - 4) + \cdots + (E_1 - 2(m-1)) = 0.$$

It follows that $m\,E_1 - (m-1)m = 0$; or equivalently, $E_1 = m - 1$. Finally,

$$\begin{aligned} E_x &= (E_1 - E_0) + (E_2 - E_1) + \cdots (E_x - E_{x-1}) \\ &= E_1 + (E_1 - 2) + \cdots + (E_1 - 2(x-1)) = x\,E_1 - (x-1)x = x(m-x). \end{aligned}$$

GRP(3). Let $S_x = E[T_x^2]$ denote the expected squared waiting time. The variance of the waiting time is $V[T_x] = E[T_x^2] - E^2[T_x] = S_x - E_x^2$. Then $S_0 = 0 = S_m$. For $1 \le x \le m-1$, note that T_x equals $1 + T_{x-1}$ with probability $1/2$, and $1 + T_{x+1}$ with probability $1/2$. Hence, first squaring and then taking expectation, we have $S_x = 1 + E_{x-1} + E_{x+1} + (S_{x-1} + S_{x+1})/2$; or equivalently, after substituting E_{x-1}, E_{x+1} and simplifying, we have $S_{x+1} - S_x = S_x - S_{x-1} + 2 - 4x(m-x)$. Next, proceeding as in the proof of GRP(2), but skipping the details, we get

$$S_x = x(m-x)[x^2 + (m-x)^2 - 2]/3 + x^2(m-x)^2.$$

After subtracting E_x^2, we get the variance of T_x, completing the proof. ■

Four paths and a cycle

Figure 1 shows the various networks on which our walk takes place. We call a finite section of the integer lattice a path. Without loss of generality, the nodes of a path are labeled $\{0, 1, 2, \ldots, m\}$. Thus a path has two end nodes 0 and m; all other nodes are interior nodes. Each interior node has outgoing arcs to both of its two neighboring nodes. But the outgoing arcs from the end nodes can vary, giving rise to five distinct networks:

1. A: Both end nodes are *absorbing*; that is, the end nodes have no outgoing arcs.
2. R: Both end nodes are *reflective*; that is, each end note leads only to its neighbor.
3. L: The two end nodes exhibit mixed features—Node m leads to Node $(m-1)$ only, but Node 0 leads to Node 1 and loops back to itself. Thus, Node m is reflective, while Node 0 is called "sticky." The network resembles a *lollipop*.
4. D: Both end nodes are sticky; that is, each end node leads to its neighbor and loops back to itself. The network resembles a *dumbbell* having a bulge on each end.
5. C: Node 0 leads to Node 1 and Node m, and Node m leads to Node $(m-1)$ and Node 0, giving rise to a *cycle*. A cyclic network, of course, is not a linear path.

Why do we study the above networks? The walk on absorbing Network A, which is the same as the gambler's ruin problem, stops as soon as one of the end nodes is

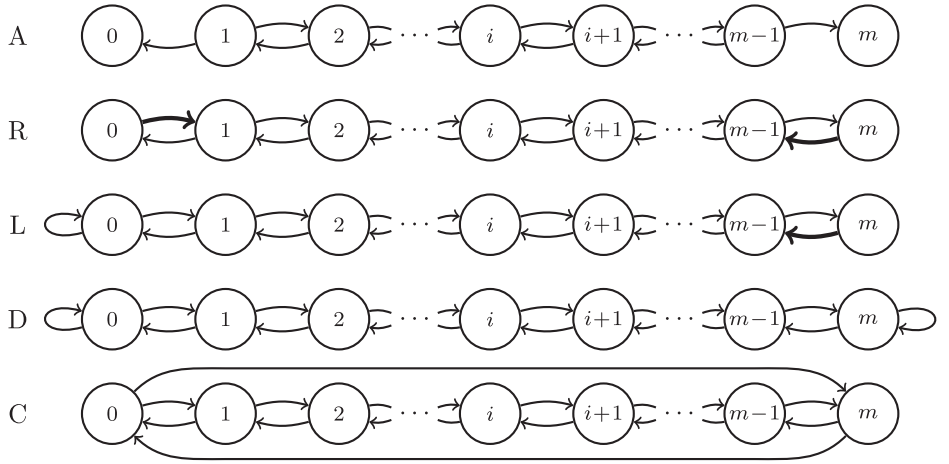


Figure 1 Linear Networks A (absorbing), R (reflective), L (lollipop), D (dumbell) with four end conditions, and cyclic Network C. A thin arrow stands for a probability of 1/2, while a thick arrow for a probability of 1.

reached. The other four networks above allow the walk to continue on and on. While our motivating problems involve only Networks R and C, we also study Networks L and D, for one can easily formulate similar motivating problems on these networks as well. Note that in Network C and Network D, each node has both in-degree and out-degree exactly 2 (a loop at a node contributes to both in-degree and out-degree of that node). Of course, lollipop Network L is intermediate between reflective Network R and dumbbell Network D.

A noteworthy connection between cyclic Network C and reflective Network R is as follows: Cyclic Network C can be obtained when the vertices of a polygon on $(m+1)$ sides are considered as nodes, and each side of the polygon is replaced by two arcs going in opposite directions. If a side of the polygon is cut off, cyclic Network C changes into reflective Network R. Also, if a vertex of the polygon is cut off (splitting the node into two nodes) then cyclic Network C again changes into reflective Network R, but with one more node. Such splitting is useful in extending the results on linear Network R to cyclic Network C, as we shall see in due time.

Of the five networks mentioned above, only cyclic Network C exhibits rotation symmetry; that is, when node labels are increased by a constant (modulo $m+1$) the network remains the same. Path Networks A, R, D and cyclic Network C exhibit a reversal symmetry; that is, the network remains the same when the node labels are reversed; that is, Node i is labeled $m-i$ for $0 \le i \le m$. However, path Network L does not have reversal symmetry, since its two end nodes exhibit dissimilar features.

Transition probabilities. Let ${}_ip_B^{(h)}$ denote the probability that the walk goes from Node i to some element of set B without hitting Node h. When $B = \{j\}$ is a singleton, we abbreviate ${}_ip_{\{j\}}^{(h)}$ by ${}_ip_j^{(h)}$. Rewriting GRP(1) of Theorem 1 using this notation, we have the following theorem.

Theorem 2. *On path Networks A, R, L, D, we have* ${}_ip_m^{(0)} = i/m$ *for* $0 \le i \le m$. *Also, by reversal symmetry, we have* ${}_{m-i}p_m^{(0)} = {}_ip_0^{(m)} = 1 - {}_ip_m^{(0)} = 1 - i/m$. *Finally, on cyclic Network C, if* $0 \le i < k \le m$, *then* ${}_ip_k^{(0)} = i/k$, *which is the same as* ${}_ip_k^{(0)}$ *on any path Network A, R, L or D with nodes* $\{0, 1, \ldots, k\}$.

Questions about walks on various networks. The walk on Network A gets absorbed (hence, stops forever) once one of the extremities is reached. Theorem 1 answers several questions regarding the walk on absorbing Network A. Here we consider walks on the other four networks R, L, D, C. We study the probability distribution in general, and the mean and the variance in particular, of the following random variables:

1. time T_R until first return to the starting position, and the number N_R of nodes (other than the starting position) visited until first return to the starting position;
2. cover time $\overline{T}$ until all nodes are visited and Node L, which is visited last; and
3. additional time ${}_L T_R$ to return to the starting position after visiting all nodes and the number of nodes ${}_L N_R$ visited during this return journey to the starting position after visiting all nodes.

Various parts of the above problems can be found in the literature. For example, three books by Ross [8–10] discuss or pose as exercises some of these questions. The asymmetric case is studied by Vallois [12]. He uses martingale techniques to obtain the probability generating function of the cover time ($\overline{T}$) on Network R as the ratio of two polynomials. Starting from an interior node on Network R, a most comprehensive study on the joint distribution of the cover time, the last vertex visited and the time taken to go from one extremum to the other is given by Chong, Cowan and Holst [2]. They express the joint pgf in terms of trigonometric functions with complex arguments or in terms of hyperbolic functions with real arguments. Sarkar [11] discusses similar problems in the context of an asymmetric random walk on the vertices of a polygon. All these references use advanced methods involving intricate application of mathematical probabilities. Except for [2], all these references restrict attention to expected times only. Here, we focus solely on the symmetric case, use elementary methods, and include the study of variance of all random variables.

Mean and variance of transit time between two nodes

Let us derive the mean (i.e., the mathematical expectation) and the variance of the time the walk takes to go from Node i to Node k on Networks R, L, D, and C. These results are needed in answering the three waiting time questions raised above. Let ${}_i T_B$ denote the waiting time the walk takes to go from Node i to any one element of set B. Let the mean, the mean square and the variance of be denoted respectively by ${}_i E_B$, ${}_i S_B$ and ${}_i V_B$. With this notation, the game duration of the gambler's ruin problem is denoted by ${}_i T_{\{0,m\}}$. When set $B = \{k\}$ is a singleton, we write ${}_i T_{\{k\}} = {}_i T_k$; etc.

Mean and variance of transit time on reflective Network R. For $0 \le i < k \le m$, the transit time ${}_i T_k$ on reflective Network R with nodes $\{0, 1, 2, \ldots, m\}$ is the same as it is on reflective Network R with nodes $\{0, 1, 2, \ldots, k\}$. This walk can be thought of as a walk on absorbing Network A with nodes $\{-k, -k+1, \ldots, -1, 0, 1, 2, \ldots, k\}$ starting from Node i, if we imagine that after visiting Node 0, if at all, the walk is equally likely to go in the positive or the negative direction. After suitable renumbering (adding k to each node number), we are essentially considering the game duration of a walk on Network A with nodes $\{0, 1, 2, \ldots, 2k\}$ starting at Node $k + i$. From GRP(2) and GRP(3) of Theorem 1, we have the following result.

Theorem 3. *On reflective Network R, for $0 \le i < k \le m$, the transit time ${}_i T_k$ has mean ${}_i E_k = k^2 - i^2 = (k - i)(k + i)$ and variance*

$$ {}_i V_k = 2(k - i)(k + i)(k^2 + i^2 - 1)/3. $$

In particular, the transit time ${}_{m-1}T_m$ *has mean* ${}_{m-1}E_m = 2m-1$ *and variance* ${}_{m-1}V_m = 4m(m-1)(2m-1)/3 = 2\binom{2m}{3}$. *Also, the transit time* ${}_0T_m$ *has mean* ${}_0E_m = m^2$ *and variance* ${}_0V_m = 2m^2(m^2-1)/3 = 4m\binom{m+1}{3}$.

As an alternate proof, since ${}_0T_m = {}_0T_1 + {}_1T_2 + \cdots + {}_{m-1}T_m$, we have

$$ {}_0E_m = \sum_{i=1}^{m} {}_{i-1}E_i = \sum_{i=1}^{m} (2i-1) = m^2. $$

And since the component random variables ${}_0T_1$, ${}_1T_2, \ldots,$ ${}_{m-1}T_m$ are *independent*, we have

$$ {}_0V_m = \sum_{i=1}^{m} {}_{i-1}V_i = 2\sum_{i=1}^{m} \binom{2i}{3} = \frac{2}{3}m^2(m^2-1) = 4m\binom{m+1}{3}. $$

Remark 1. For $0 \le k < i \le m$, the transit time ${}_iT_k$ on Network R has the same distribution as the transit time ${}_{(m-i)}T_{(m-k)}$ on Network R, seen by reverse numbering the nodes. Hence, by Theorem 3, the transit time ${}_iT_k$ has mean $(i-k)(2m-i-k)$ and variance $2(i-k)(2m-i-k)[(m-k)^2+(m-i)^2-1]/3$. In particular, the transit time ${}_1T_0$ has mean ${}_1E_0 = 2m-1$ and variance ${}_1V_0 = 2\binom{2m}{3}$.

Mean and variance of transit time on lollipop Network L. The walk on lollipop Network L differs from that on Network R only because the waiting time at Node 0, before the walk moves to Node 1, is given by a geometric random variable of probability $p = 1/2$. Therefore, let us imagine further that the sticky barrier 0 is replaced by two copies of 0 (denoted by -0 and $+0$); and the walk takes place on the Network A with nodes $\{-k, -k+1, \ldots, -1, -0, +0, 1, 2, \ldots, k\}$. The key idea is that the time spent bouncing between -0 and $+0$ will have the requisite geometric distribution of time spent at Node 0 by a walk on Network L. By further renumbering of the nodes, the walk reduces to a GRP on $\{0, 1, \ldots, 2k+1\}$ starting from Node $(k+i+1)$. From GRP(2) and GRP(3) of Theorem 1, we have the following result.

Theorem 4. *On lollipop Network L, for* $0 \le i < k \le m$, *the transit time* ${}_iT_k$ *has mean* ${}_iE_k = (k-i)(k+i+1)$ *and variance*

$$ {}_iV_k = (k-i)(k+i+1)[(k-i)^2+(k+i+1)^2-2]/3. $$

In particular, the transit time ${}_{m-1}T_m$ *has mean* ${}_{m-1}E_m = 2m$ *and variance* ${}_{m-1}V_m = 2m(4m^2-1)/3 = 2\binom{2m+1}{3}$. *Also, the transit time* ${}_0T_m$ *has mean* ${}_0E_m = m(m+1)$ *and variance* ${}_0V_m = m(m+1)(2m^2+2m-1)/3$.

Without applying Theorem 1, since ${}_0T_m = {}_0T_1 + {}_1T_2 + \cdots + {}_{m-1}T_m$, we have

$$ {}_0E_m = \sum_{i=1}^{m} {}_{i-1}E_i = \sum_{i=1}^{m} (2i) = m(m+1). $$

It follows that we have

$$ {}_0V_m = \sum_{i=1}^{m} {}_{i-1}V_i = 2\sum_{i=1}^{m} \binom{2i+1}{3} = m(m+1)(2m^2+2m-1)/3. $$

Remark 2. For $0 \le k < i \le m$, by utilizing reverse numbering, the transit time ${}_iT_k$ on lollipop Network L has the same distribution as the transit time ${}_{(m-i)}T_{(m-k)}$ on reflecting Network R, whose mean and variance are given in Remark 1.

Mean and variance of transit time on dumbbell Network D. For $0 \le i < k \le m$, the transit time ${}_iT_k$ on dumbbell Network D has the same distribution as the transit time ${}_iT_k$ on lollipop Network L. But for $0 \le k < i \le m$, by utilizing reverse numbering, the transit time ${}_iT_k$ on dumbbell Network D has the same distribution as the transit time ${}_{(m-i)}T_{(m-k)}$ on lollipop Network L. Hence, the next result follows from Theorem 4.

Theorem 5. *On dumbbell Network D, for $0 \le i < k \le m$, the transit time ${}_iT_k$ has mean ${}_iE_k = (k-i)(k+i+1)$ and variance*

$$ {}_iV_k = (k-i)(k+i+1)[(k-i)^2 + (k+i+1)^2 - 2]/3. $$

But for $0 \le k < i \le m$, the transit time ${}_iT_k$ has mean $(i-k)(2m-k-i+1)$ and variance

$$ {}_iV_k = (i-k)(2m-k-i+1)[(i-k)^2 + (2m-k-i+1)^2 - 2]/3. $$

In particular, the transit time ${}_{m-1}T_m$, as well as ${}_1T_0$, has mean $2m$ and variance $2\binom{2m+1}{3}$. Also, the transit time ${}_0T_m$, as well as ${}_mT_0$, has mean $m(m+1)$ and variance $m(m+1)(2m^2+2m-1)/3$.

Mean and variance of transit time on cyclic Network C By rotation symmetry of cyclic Network C, it suffices to study the transit time ${}_iT_0$ from Node i (with $i \neq 0$) to Node 0. Indeed, ${}_iT_0$ on cyclic Network C behaves as the game duration of a GRP on $\{0, 1, 2, \ldots, m+1\}$ starting at Node i. Here, Node 0 of cyclic Network C has been split to form two nodes 0 and $m+1$ of absorbing Network A (with $m+2$ nodes in all). Hence, from GRP(2) and GRP(3) of Theorem 1, we have the following result.

Theorem 6. *On cyclic Network C, starting from Node i (with $i \neq 0$), the waiting time until the walk reaches Node 0 has mean $i(m+1-i)$ and variance*

$$ i(m+1-i)[i^2 + (m+1-i)^2 - 2]/3. $$

In particular, starting from Node 1, the waiting time to reach Node 0 has mean m and variance $2\binom{m+1}{3}$.

First return to the starting position

On each Network R, L, D, and C, let us follow the walk from its starting node until it leaves that node and then returns to it for the first time. We answer two questions: (1) How many steps are taken until the walk returns to the starting position? (2) How many nodes have been visited by the time the walk returns to the starting position?

Time of first return to the starting position On linear Networks R, L, D, we consider separately three cases according as the starting position is (1) Node 0, (2) some interior node, or (3) Node m.

Case (1): Starting position is Node 0. On reflective Network R, in the first step the walk surely reaches Node 1. Thereafter, we wait until it returns to Node 0. Hence, $T_R = 1 + {}_1T_0$. Therefore, by Remark 1, on Network R, T_R has mean $1 + (2m-1) = 2m$ and variance $2\binom{2m}{3}$. Similarly, on lollipop Network L, after a waiting time G at Node 0 that is distributed geometrically with $p = 1/2$, the walk reaches Node 1. Hence, $T_R = G + {}_1T_0$, where G and ${}_1T_0$ are independent. Hence, by Remark 2, on Network L, T_R has mean $2 + (2m-1) = 2m+1$ and variance $2 + 2\binom{2m}{3}$. Finally, on dumbbell Network D, the mean and the variance of $T_R = G + {}_1T_0$, in view of Theorem 5, are $2 + 2m$ and $2 + 2\binom{2m+1}{3}$ respectively.

Case (2): Starting position is an interior Node i $(1 \le i \le m-1)$. In one step the walk reaches Node $(i-1)$ with probability $1/2$, or Node $(i+1)$ with probability $1/2$. Note that the return to Node i involves shorter networks (of possibly different types) on the two sides of Node i. Specifically, with probability $1/2$, the return time T_R on Network R, L or D, behaves like $1 + {}_{i-1}T_i$ on Network R, L or L, respectively; and with probability $1/2$, it behaves like $1 + {}_{m-i-1}T_{m-i}$ on Network R, R or L, respectively. Next, using Theorems 3 and 4, we obtain the mean and variance of $1 + {}_{i-1}T_i$ and $1 + {}_{m-i-1}T_{m-i}$ on networks R and L. Thereafter, by the law of total expectation, the mean return time on Networks R, L, and D are respectively m, $m + 1/2$, and $m + 1$. Also, by the law of total variance, the variance of the return time on Networks R, L, and D are, respectively,

$$
\begin{aligned}
V(R) &= \binom{2i}{3} + \binom{2(m-i)}{3} + (m-2i)^2 = 2\binom{m}{3} + m(m-2i)^2, \\
V(L) &= \binom{2i+1}{3} + \binom{2(m-i)}{3} + (m-2i-1/2)^2 \\
&= 2\binom{m+1/2}{3} + (m+1/2)(m-2i-1/2)^2, \text{ and} \\
V(D) &= \binom{2i+1}{3} + \binom{2(m-i)+1}{3} + (m-2i)^2 \\
&= 2\binom{m+1}{3} + (m+1)(m-2i)^2.
\end{aligned}
$$

Case (3): Starting position is Node m. Then on Networks R and D, in view of reversal symmetry, the distribution of T_R is the same as it is when the starting position is Node 0. That is, on reflective Network R, T_R has mean $2m$ and variance $2\binom{2m}{3}$; and on dumbbell Network D, T_R has mean $2 + 2m$ and variance $2 + 2\binom{2m+1}{3}$. However, on lollipop Network L, which does not exhibit reversal symmetry, the return time $T_R = 1 + {}_{m-1}T_m$ has mean $2m + 1$ and variance $2\binom{2m+1}{3}$, using Theorem 4.

Remark 3. On cyclic Network C, the return time ${}_iT_i$ is the same for all $0 \le i \le m$, and each one has the same distribution as $1 + {}_1T_0$ on cyclic Network C. Hence, by Theorem 6, ${}_iT_i$ has mean $1 + m$ and variance $2\binom{m+1}{3}$.

Number of nodes visited until first return to the starting position. Suppose that the random walk starts from Node i. By the time the walk returns to Node i, it surely has visited at least one node other than Node i. How many nodes altogether (other than Node i) are visited before the walk returns to Node i? Let ${}_iN_i$ denote the number of nodes visited starting from Node i until the walk returns to Node i, not counting Node i itself.

The distribution of ${}_iN_i$ is the same on all three linear Networks R, L and D, since visit count will be insensitive to whether the end nodes of the path are reflective or sticky. Again, we consider the three cases according as the starting position is (1) Node 0, (2) some interior node i for $1 \le i \le m-1$, or (3) Node m.

Case (1): Starting position is Node 0. In one step (or after a waiting time that is geometrically distributed with $p = 1/2$) the walk reaches Node 1. Let k be the node farthest from Node 0 visited by the walk before it returns to Node 0. If $1 \le k \le m-1$,

the walk has moved from Node 1 to Node k without hitting Node 0, which happens with probability ${}_1p_k^{(0)}$. Thereafter it has gone from Node k to Node 0 without hitting Node $(k+1)$, which happens with probability ${}_kp_0^{(k+1)}$. Hence, using Theorem 2, for $1 \le k \le m-1$, we have

$$P[{}_0N_0 = k] = {}_1p_k^{(0)} \cdot {}_kp_0^{(k+1)} = \frac{1}{k} \cdot \frac{1}{k+1} = \frac{1}{k} - \frac{1}{k+1}.$$

For $k = m$, the walk must move from Node 1 to Node m without hitting Node 0, which happens with probability ${}_1p_m^{(0)}$. Hence, $P[{}_0N_0 = m] = {}_1p_m^{(0)} = 1/m$.

Remark 4. On networks R, L and D, for $1 \le k \le m-1$, we have $P[{}_0N_0 = k] = \frac{1}{k} - \frac{1}{k+1} > 0$; and $P[{}_0N_0 = m] = \frac{1}{m} > 0$. Hence,

$$\sum_{k=1}^{m} P[{}_0N_0 = k] = \sum_{k=1}^{m-1} \left(\frac{1}{k} - \frac{1}{k+1} \right) + \frac{1}{m} = 1,$$

which shows that we have a genuine probability distribution of ${}_0N_0$. Moreover, this is a discrete version of the Zipf(2) distribution, and a distribution like this with a heavy tail tends to have an unexpectedly large variance. Indeed, the mean of ${}_0N_0$ is

$$H_m = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m},$$

the mean square of ${}_0N_0$ is $2m - H_m$, and the variance is $2m - H_m(1 + H_m)$.

Case (2): Starting position is some interior Node i (for $1 \le i \le m-1$). Starting from Node i, the walk goes to Node $(i-1)$ or Node $(i+1)$ with probability $1/2$ each. The return to Node i involves shorter networks. Specifically, (by renumbering the nodes suitably) ${}_iN_i$ equals ${}_0N_0$ on Network A with nodes $\{0, 1, 2, \ldots, i\}$ or ${}_0N_0$ on Network A with nodes $\{0, 1, 2, \ldots, m-i\}$ with probability $1/2$ each. Since ${}_iN_i$ has a finite support, writing the probability mass function as a vector $(P({}_iN_i = k), 1 \le k \le m)$, by the law of total probability, we evaluate it to be

$$\frac{1}{2}\left(\frac{1}{1\cdot 2}, \frac{1}{2\cdot 3}, \ldots, \frac{1}{(i-1)\cdot i}, \frac{1}{i}, 0, \ldots, 0 \right)$$
$$+ \frac{1}{2}\left(\frac{1}{1\cdot 2}, \frac{1}{2\cdot 3}, \ldots, \frac{1}{(m-i-1)\cdot (m-i)}, \frac{1}{m-i}, 0, \ldots, 0 \right).$$

Consequently, by the law of total expectation, $E({}_iN_i) = (H_i + H_{m-i})/2$, and by squaring ${}_iN_i$ and then applying the law of total expectation, we have

$$E({}_iN_i^2) = [2i - H_i + 2(m-i) - H_{m-i}]/2 = m - [H_i + H_{m-i}]/2.$$

Hence, the variance of ${}_iN_i$ is

$$V({}_iN_i) = E({}_iN_i^2) - \left(\frac{H_i + H_{m-i}}{2} \right)^2 = m - \frac{H_i + H_{m-i}}{2} \left(\frac{H_i + H_{m-i}}{2} + 1 \right).$$

Note that both $E({}_iN_i)$ and $V({}_iN_i)$ are invariant when i is replaced by $m-i$, as anticipated from reversal symmetry.

Case (3): Starting position is Node m. Since the feature of the end node is irrelevant for the distribution of ${}_mN_m$, it is the same as that of ${}_0N_0$ given in Remark 4.

Finally, the distribution of ${}_iN_i$ on cyclic Network C, by rotation symmetry, is the same for all Nodes i. Without loss of generality, let us assume that the walk starts at Node 0 and goes to Node 1 at time $t = 1$. Let us split Node 0 of cyclic Network C into Nodes 0 and $m + 1$, thereby forming absorbing Network A on $m + 2$ nodes, but do keep in mind that the new Node $m + 1$ is essentially the same as the new Node 0. Next, let ${}_1N_e$ denote the number of nodes visited until the walk on absorbing Network A starting from Node 1 reaches either extremity (where we count Node 1, but not the extremity). Then the distribution of ${}_0N_0$ on cyclic Network C is the same as the distribution of ${}_1N_e$ on any path Network with nodes $\{0, 1, 2, \ldots, m + 1\}$ (with the understanding that Nodes 0 and $m + 1$ are one and the same), or the distribution of ${}_0N_0$ on any path Network with nodes $\{0, 1, 2, \ldots, m\}$, given in Remark 4.

Visiting all nodes

On linear networks R, L, D, and cyclic Network C, how many steps are taken by the walk until all nodes are visited? Also, which node is visited last; that is, where is the walk when all nodes are visited? Let ${}_i\overline{T}$ denote the first time $t > 0$ by when all nodes are visited, starting from Node i at time $t = 0$; and let $L = X({}_i\overline{T})$ denote the node visited last; that is, where the walk is at the instant all nodes are visited.

Time to visit all nodes. Suppose that the walk starts at Node i, where $0 \le i \le m$. Then the cover time on each of the linear networks R, L, and D can be written as the sum of two independent random variables: Either ${}_i\overline{T} = {}_iT_{\{0,m\}} + {}_mT_0$ with probability ${}_ip_m^{(0)} = i/m$, or ${}_i\overline{T} = {}_iT_{\{0,m\}} + {}_0T_m$ with probability $(1 - i/m)$. Furthermore, Networks R and D have reversal symmetry, which implies that ${}_mT_0 = {}_0T_m$. Therefore, on reflective Network R, by Theorems 1 and 3, the mean cover time is $i(m - i) + m^2$ and the variance is

$$i(m - i)[(m - i)^2 + i^2 - 2]/3 + 2m^2(m^2 - 1)/3.$$

Likewise, on dumbbell Network D, by Theorems 1 and 5, the mean cover time is $\{i(m - i) + m^2\} + m$ and the variance is

$$i(m - i)[(m - i)^2 + i^2 - 2]/3 + m(m + 1)(2m^2 + 2m - 1)/3.$$

However, lollipop Network L does not have reversal symmetry. On lollipop Network L, which is intermediate between networks R and D, using the law of total expectation, Theorems 1 and 4 and Remark 2, the mean cover time equals the weighted average of mean cover times on networks R and D with weights ${}_ip_0^{(m)}$ and ${}_ip_m^{(0)}$, respectively, simplifying to

$$\begin{aligned}[i(m - i) + m^2]\,{}_ip_0^{(m)} + [i(m - i) + m^2 + m]\,{}_ip_m^{(0)} \\ = [i(m - i) + m^2] + m\,{}_ip_m^{(0)} = [i(m - i) + m^2] + (m - i) \\ = m^2 + (i + 1)(m - i).\end{aligned}$$

Likewise, using the law of total variance, Theorems 1 and 4 and Remark 2, the variance of the cover time simplifies to

$$\begin{aligned}i(m - i)[(m - i)^2 + i^2 - 2]/3 + 2m^2(m^2 - 1)/3 \\ + i(m - i) + (m - i)(m + 1)(4m - 1)/3.\end{aligned}$$

On cyclic network C, the cover time is independent of the starting position i. Assume without loss of generality that the walk starts at Node 1. We write ${}_1\overline{T}$ as the sum of *independent* components that record the time until one additional node is visited for the first time. At each of these one-step advancement epochs, the nodes visited so far form a connected chain of length l, say, where $1 \le l \le m$. Renumber the newest visited node as 1, its nearest unvisited node as 0, the node at the opposite end of the connected-chain-of-visited-nodes as l and the unvisited node next to l as $(l+1)$, with the understanding that Node $m+1$ is the same as Node 0. This forms a network A on nodes $\{0, 1, \ldots, l+1\}$. Thus

$$ {}_1\overline{T} = {}_1T_{\{0,\,2\}} + {}_1T_{\{0,\,3\}} + \ldots + {}_1T_{\{0,\,m+1\}}. \tag{1} $$

By Theorem 1, the cover time ${}_1\overline{T}$ has mean $\sum_{j=1}^{m} j = m(m+1)/2$. Also, since the components on the right hand side of equation 1 are independent, ${}_1\overline{T}$ has variance $\sum_{j=1}^{m} 2\binom{j+1}{3} = 2\binom{m+2}{4}$.

Remark 5. By the central limit theorem, the cover time of cyclic Network C with nodes $\{0, 1, \ldots, m\}$ by a walk starting at any node i with $0 \le i \le m$ is asymptotically normal; that is, for large m, the distribution of $\left[{}_i\overline{T} - \binom{m+1}{2}\right] / \sqrt{2\binom{m+2}{4}}$ is approximately standard normal.

The last node visited. On linear Networks R, L. and D, if the walk starts at Node i $(0 \le i \le m)$, then the last node to be visited is Node 0 if Node m was visited before Node 0, which happens with probability i/m in view of Theorem 2. Otherwise, the last node is Node m with probability $(1 - i/m)$.

On cyclic Network C, there is a pleasantly surprising result regarding the distribution of the last node visited, which many people find counter-intuitive! The result, with proof, appears in Ross [**8**]. We reproduce it below.

Theorem 7. *On cyclic Network C, irrespective of the starting Node i, every node $j \ne i$ is equally likely to be the last node visited.*

Proof. If starting from Node i the walk visits Node j last, then the second last node must be either Node $(j-1)$ or Node $(j+1)$, the neighboring nodes of j. First, suppose that Node $(j-1)$ is the second last node. Then starting from Node i the walk has visited $(j+1)$ before $(j-1)$ and j. This happens with probability Q_{ij} (which we don't need to evaluate). Thereafter, the walk has gone from $(j+1)$ to $(j-1)$ without hitting j. What is the conditional probability of this happening? Renumber Node j as 0, $(j+1)$ as 1, and $(j-1)$ as m going clockwise; see Figure 2(a). Then, the above conditional probability is nothing but ${}_1p_m^{(0)} = 1/m$, by Theorem 2.

Next, suppose that Node $(j+1)$ is the second last node. In this case, the walk has visited $(j-1)$ before $(j+1)$ and j (happening with probability $1 - Q_{ij}$), then the walk has gone from $(j-1)$ to $(j+1)$ before hitting j. This conditional probability is also ${}_1p_m^{(0)} = 1/m$, seen by renumbering j as 0, $(j-1)$ as 1 and $(j+1)$ as m going counterclockwise; see Figure 2(b). Hence,

$$ P[L = j] = Q_{ij}(1/m) + (1 - Q_{ij})(1/m) = 1/m $$

is the same for all $j \ne i$, and is independent of i. This completes the proof. ∎

Remark 6. In the above proof we did not have to evaluate Q_{ij}. But we have all the ingredients to evaluate it. Indeed, $Q_{ij} = {}_ip_{j+1}^{(j-1)} = {}_{j-i-1}p_{m-1}^{(0)}$. [The last equality is justified by renumbering the nodes counterclockwise so that Node $(j-1)$ becomes 0,

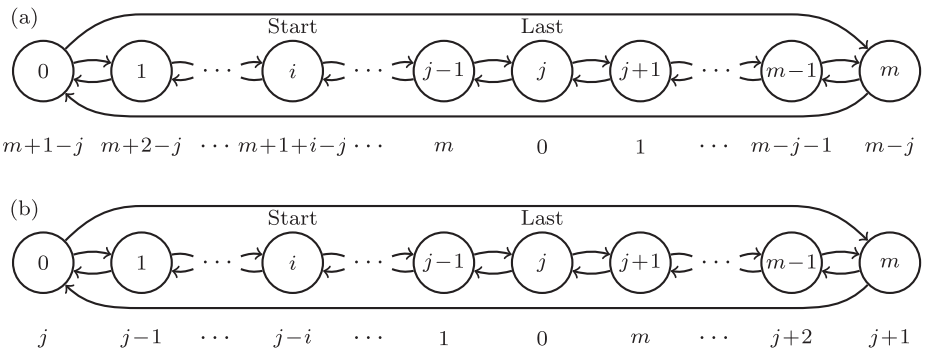


Figure 2 Renumbering the nodes of cyclic Network C, if starting from Node i, the walk visits Node j last, according as which node is visited second last. In a), if the walk visits Node $j+1$ before Nodes $j-1$ and j, then renumber the nodes clockwise so that Node j becomes Node 0. In b), if the walk visits Node $j-1$ before Nodes $j+1$ and j, then renumber nodes counterclockwise so that Node j becomes Node 0.

Node i becomes $(j-i-1)$, and Node $(j+1)$ becomes $(m-1)$.] Then by Theorem 2, we have $Q_{ij} = (j-i-1)/(m-1)$.

Return to the starting position after visiting all nodes

We follow the walk from the starting Node i until it has visited all nodes at least once and is presently at Node $L = X(\,{}_i\overline{T}) \neq i$. As we studied in the previous section, L is a random node. Let us continue to watch the walk until it returns to the starting Node i. How much longer will it take to return to the starting Node i after visiting all nodes? How many nodes will be visited during the interim?

Time to return to the starting position after visiting all nodes. Suppose that the walk starts from Node i at time $t = 0$, and it visits Node L last. Let us denote by ${}_L T_i$ the additional time to return to Node i after visiting all nodes. As described in the previous section, on networks R, L and D, the last node visited is Node 0 (with probability i/m) or Node m (with probability $1 - i/m$). Therefore, ${}_L T_i$ equals ${}_0 T_i$ with probability i/m, or it equals ${}_m T_i$ with probability $1 - i/m$.

Using the law of total expectation, by Theorem 3 and Remark 1, on reflective Network R, ${}_L T_i$ has mean

$$E(\mathrm{R}) = i^2(i/m) + (m-i)^2(1-i/m) = m^2 - 3mi + 3i^2.$$

Next, using the law of total variance, by Theorem 3 and Remark 1, on reflective Network R, ${}_L T_i$ has variance

$$\begin{aligned} V(\mathrm{R}) =& \{[2i^2(i^2-1)/3]\,(i/m) + [2(m-i)^2[(m-i)-1]]\,(1-i/m)\} \\ &+ [i^4\,(i/m) + (m-i)^4\,(1-i/m)] - (m^2 - 3mi + 3i^2)^2, \end{aligned}$$

which, after simplification, becomes

$$V(\mathrm{R}) = 2m^2[(m-2i)^2 - 1]/3 + i(m-i)[(m-i)^2 + i^2 + 6]/3.$$

As anticipated from reversal symmetry, both $E(\mathrm{R})$ and $V(\mathrm{R})$ are invariant when i is replaced by $m-i$.

Likewise, on lollipop Network L, by Theorem 4 and Remark 2, ${}_LT_i$ has mean

$$E(\mathrm{L}) = i(i+1)(i/m) + (m-i)^2(1-i/m) = E(\mathrm{R}) + i^2/m;$$

and the variance, after algebraic simplification, is equal to

$$\begin{aligned} V(\mathrm{L}) &= V(\mathrm{R}) + \frac{i^2}{3m}\left[-\left(8+\frac{3}{m}\right)i^2 + 6(3m+1)i - (6m^2+1)\right] \\ &= V(\mathrm{R}) + g(i). \end{aligned}$$

The function $g(i)$ is increasing on $0 \le i \le m$; it is negative for sufficiently small i, and positive for large $i \le m$. For example, if $m = 20$, then the variance of ${}_LT_i$ is smaller on lollipop Network L than on reflective Network R for $0 \le i \le 7$, and larger for $8 \le i \le 20$.

On dumbbell Network D, ${}_LT_i$ has mean

$$E(\mathrm{D}) = [i^2(i+1) + (m-i)^2(m-i+1)]/m = E(\mathrm{R}) + i^2/m + (m-i)^2/m,$$

and, after simplification, has variance

$$V(\mathrm{D}) = V(\mathrm{R}) + g(i) + g(m-i).$$

Note that $g(i) + g(m-i)$ is positive for all $1 \le i \le m$. Again, as anticipated from reversal symmetry, both $E(\mathrm{D})$ and $V(\mathrm{D})$ are invariant when i is replaced by $m-i$.

On cyclic Network C, recall from Theorem 7 that the last node visited is *equally likely* to be any node other than the starting position (which we will assume without loss of generality is Node 0). Also recall from Theorem 6 that the transit time from any Node i to Node 0 is the same as the game duration of a gambler's ruin problem on nodes $\{0, 1, \ldots, m+1\}$ starting from Node i; that is, ${}_iT_0(\mathrm{C}) = {}_iT_{\{0,\,m+1\}}(\mathrm{A})$. Using the law of total expectation and GRP(2) of Theorem 1, the mean of the return time ${}_LT_0$ is

$$E(\mathrm{C}) = \frac{1}{m}\sum_{i=1}^{m} i(m+1-i) = \frac{1}{m}\binom{m+2}{3} = (m+2)(m+1)/6.$$

Likewise, using the law of total variance and GRP(3) of Theorem 1, after some simplifications, we find the variance of the return time ${}_LT_0$ to be

$$\begin{aligned} V(\mathrm{C}) &= \frac{1}{m}\sum_{i=1}^{m} i(m+1-i)\{i^2 + (m+1-i)^2 - 2\}/3 \\ &\qquad + \frac{1}{m}\sum_{i=1}^{m} i^2(m+1-i)^2 - \left\{\frac{1}{m}\sum_{i=1}^{m} i(m+1-i)\right\}^2 \\ &= \frac{1}{m}\binom{m+2}{3}\left\{\frac{2}{3}\left(\frac{3(m+1)^2-2}{10}\right) - \frac{2}{3} + \frac{(m+1)^2+1}{5} + \frac{1}{m}\binom{m+2}{3}\right\} \\ &= (m+2)(m+1)(m-1)(7m+16)/180. \end{aligned}$$

Number of nodes visited during return to the starting position after visiting all nodes. Suppose that the walk starts from Node i and it visits Node L last. Define ${}_LN_i$ to be the number of nodes visited during the return trip to the starting position after visiting all nodes (counting Node L, but not counting the starting Node i). On linear

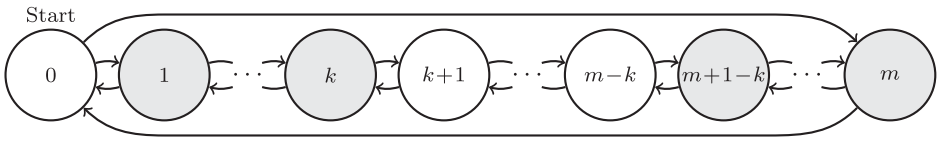


Figure 3 On cyclic Network C, $\{{}_L N_0 = k\}$ only if the last node is among Nodes 1 through k or Nodes $(m+1-k)$ through m (shown with nodes shaded).

Networks R, L, and D, ${}_L N_i$ equals i with probability i/m, or $(m-i)$ with probability $(1-i/m)$. Hence, upon simplification, the number of nodes visited during the return to the starting position has mean $[i^2+(m-i)^2]/m = m-2i+2i^2/m$ and variance

$$[i^3+(m-i)^3]/m - \left(m-2i+2i^2/m\right)^2 = i(m-i)(m-2i)^2/m^2.$$

On cyclic Network C, without loss of generality the starting position is Node 0. After visiting all vertices, the walk returns to 0 either from Node 1 or from Node m. For $\{{}_L N_0 = k\}$ to hold, the last node visited must have been in either $\{1, 2, \ldots, k\}$ in the former case, or in $\{m, m-1, \ldots, m+1-k\}$ in the latter case, with probability $1/m$ for each of these possible nodes by Theorem 7. See Figure 3.

It is enough to double the probability in the former case, since in the latter case we can renumber the nodes counterclockwise (keeping the same number for Node 0, and changing Node h to $m+1-h$) and then imitate the former case. Suppose that $\{{}_L N_0 = k\}$ and the last node visited was j $(1 \le j \le k)$. Then the walk must have moved from Node $L = j$ to k without hitting 0, and from Node k to 0 without hitting $(k+1)$. So, for all $k = 1, 2, \ldots, m$, we have

$$Pr({}_L N_0 = k) = 2\sum_{j=1}^{k} \frac{1}{m}\, {}_j p_k^{(0)}\, {}_k p_0^{(k+1)} = 2\sum_{j=1}^{k} \frac{1}{m}\cdot\frac{j}{k}\cdot\frac{1}{k+1} = \frac{1}{m}.$$

Thus, we have proved that ${}_L N_0$, the number of nodes visited during the return trip from the last visited node L to the starting Node 0, is a discrete uniform random variable. This is another pleasantly surprising result that does not seem to have been documented in the literature!

Theorem 8. *On the cyclic Network C, starting from any Node i, if the last node visited is Node $L \neq i$, then during the return from Node L to Node i the number of nodes visited is uniformly distributed over* $\{1, 2, \ldots, m\}$. *Hence, the mean is* $(m+1)/2$ *and the variance is* $(m^2-1)/12$.

Comparison of the linear networks

Let us summarize the effect of changing the linear path network from R to L. Indeed, this change has absolutely no effect on the last node $L = X(\overline{T})$, or on the two number-of-nodes random variables N_R and ${}_L N_R$. But the waiting-time random variables differ in distributions as follows:

1. The transit time ${}_i T_k (0 \le i < k \le m)$ increases on average by $(k-i)$; while the variance increases by $(k-i)[3(k+i)(k+i+1)+(k-i)^2-1]/3$. In particular, the transit time ${}_{m-1} T_m$ increases on average by 1; while the variance increases by $2m(2m-1)$. Also, the transit time ${}_0 T_m$ increases on average by m; while the variance increases by $m(m+1)(4m-1)/3$.

2. The time of first return to the starting position, ${}_iT_i$, increases on average by 1, 1/2, 1 in the three cases $i = 0,\ 1 \le i \le m-1,\ i = m$ respectively; while the variance changes by 2, $2i^2 + i + 1/4 - m$, and $2m(2m-1)$, respectively.
3. The cover time ${}_i\overline{T}$ to visit all nodes starting from Node i increases on average by $(m-i)$; while the variance increases by $(m-i)[i + (m+1)(4m-1)/3]$.
4. The time ${}_LT_i$ to return to the starting position i after visiting all nodes increases on average by i^2/m; while the variance changes by

$$g(i) = \frac{i^2}{3m}\left[-\left(8 + \frac{3}{m}\right)i^2 + 6(3m+1)i - (6m^2+1)\right].$$

We leave it to the reader to study the effect of changing the linear Network from L to D, or from R to D, on the transit time between two nodes, return time to the starting position, cover time and return time to the starting position after visiting all nodes.

Solutions to the motivating problems

Carnival elevator problem solved. The $(G+10)$-story haunted high-rise tour will take you $2[1 +_1 T_0(\mathrm{R})]$ minutes. By Remark 1, it has a mean of $2(1+19) = 40$ minutes and a standard deviation of $2\sqrt{240} \approx 31$ minutes. During the tour you will have visited ${}_0N_0$ additional floors other than Floor G, with a distribution given by Remark 4: It takes on values $\{1, 2, \ldots, 10\}$ with probability mass function

$$(1/2, 1/6, 1/12, 1/20, 1/30, 1/42, 1/56, 1/72, 1/90, 1/10).$$

It has a mean of $H_{10} = 2.93$ and a standard deviation of $\sqrt{2(10) - H_{10}(1+H_{10})} = 2.91$. In any one tour the probability of visiting the topmost tenth floor is $1/10$. So the number of tours until you visit the topmost tenth floor is a geometrically distributed with $p = 1/10$ with a mean of 10 and a standard deviation of $\sqrt{90} \approx 9.5$.

Carnival rotating multiplex problem solved. The duration you will spend in the six-theater multiplex starting from Theater 0 until first return to Theater 0 is $5[T_R(\mathrm{C})]$ minutes, with a mean of 30 minutes and standard deviation of $5\sqrt{40} = 31.6$ minutes. By that time you would have seen $N_R(\mathrm{C})$ shows, with a distribution given by Remark 4; it has support $\{1, \ldots, 5\}$ and probability mass function $(1/2, 1/6, 1/12, 1/20, 1/5)$; hence, the mean is $H_5 = 137/60 = 2.28$ and standard deviation is $\sqrt{2(5) - H_5(1+H_5)} = 1.72$.

To see all 6 performances in the multiplex you will need $5[\overline{T}(\mathrm{C})]$ minutes, which, by the statement after equation (1), has a mean of $5(15) = 75$ minutes and a standard deviation of $5\sqrt{70} = 42$ minutes. By, Theorem 7, the theater you will see last is equally likely to be any one of $1, \ldots, 5$.

After receiving your complimentary hat you will wait $5[{}_LT_0(\mathrm{C})]$ minutes to return to Theater 0. This waiting time has a mean of $5(7) = 35$ minutes and a standard deviation of $5\sqrt{47.6} = 34.5$ minutes. After receiving your hat you are equally likely to watch $1, \ldots, 5$ distinct shows.

However, if you know that you received your hat in Theater i, where $1 \le i \le 5$, then the waiting time to return to Theater 0 will be $5[{}_iT_{\{0,6\}}(\mathrm{R})]$ minutes. By Theorem 6, this waiting time has a mean of $5i(6-i)$ minutes and a standard deviation of $5\sqrt{i(6-i)\{i^2 + (6-i)^2 - 2\}/3}$ minutes. Also, after receiving your hat in Theater i until you face Theater 0, you will see ${}_iN_0$ distinct theaters, whose support is $\{1, \ldots, 5\}$ with probability mass function given by $(1/2, 1/6, 1/12, 1/20, 1/5)$ if

$i = 1, 5$; $(0, 1/3, 1/6, 3/10, 2/10)$ if $i = 2, 4$; and $(0, 0, 1/2, 3/10, 2/10)$ if $i = 3$. Since L is uniformly distributed over $\{1, \ldots, 5\}$, by taking a simple average of these five probability mass functions, we can verify that ${}_L N_0$ is also uniformly distributed over $\{1, \ldots, 5\}$, as claimed in Theorem 8.

Asymptotic location of the walk

What if we let the walk continue on and on? Where will the walk be in the long run? There is a general result for the long-run stationary distribution of the location of an *aperiodic* (defined in the next two sentences) walk on a connected, finite graph. (The period of a walk is the greatest common divisor τ of all times when the walk can possibly return to the starting node. A walk is called aperiodic if the period is one.) To describe the stationary distribution on any graph, first replace each edge by two arcs—one in each possible direction—and replace a loop at a node by an arc from that node to itself. Thereafter, record the in-degree of each node; that is, the number of arcs entering into that node. Clearly, the total of in-degrees of all nodes equals the number of arcs. Then impose a uniform distribution on all arcs. In particular, this means that if we record for each transition the arc on which the walk is traveling, then in the long run the proportion of times each arc will be traveled on equals the reciprocal of the total number of arcs. Consequently, for an aperiodic walk on a connected finite graph, the long run probability that the walk will be found in any particular node is proportional to the in-degree of that node. See Lovasz [7] for details.

On linear Network L or D with any number of nodes, or on cyclic Network C with an odd number of nodes, the walk is aperiodic. On Network L, in the long run the walk is at Node m with probability $1/(2m + 1)$ and at Node i $(0 \le i \le m - 1)$ with probability $2/(2m + 1)$. On Network D, the walk is at Node i $(0 \le i \le m)$ with probability $1/(m + 1)$ each. On Network C with an odd number of nodes, the walk is at Node i $(0 \le i \le 2n)$ with probability $1/(2n + 1)$ each.

But on cyclic Network C with an even number of nodes or on linear Network R, the walk is periodic with period $\tau = 2$. The stationary distribution for a periodic walk with period τ must be described separately for each residue class of time indexed by t (modulo τ). On Network C with an even number of nodes, labeled $\{0, 1, \ldots, 2n - 1\}$, as $t \to \infty$, the location of the walk at even time $X(2t)$ is equally likely to be any one of the n vertices that are at even number of steps away from the starting position, and the location of the walk at odd time $X(2t + 1)$ is equally likely to be at any one of the other n vertices that are at odd number of steps away from the starting position.

For linear Network R with nodes $\{0, 1, \ldots, m\}$, let us assign probability masses $q_0 = 1/m = q_m$ on the two extreme nodes, and $q_j = 2/m$ for $1 \le j \le m - 1$ on the $(m - 1)$ interior nodes. Let us separate the nodes into two sets: an even set $B_0 = \{0, 2, 4, \ldots\}$ and an odd set $B_1 = \{1, 3, 5, \ldots\}$. Define $Q_0 = (q_0, q_2, q_4, \ldots)$ and $Q_1 = (q_1, q_3, q_5, \ldots)$. Note that $q_0 + q_2 + q_4 + \cdots = 1$ and $q_1 + q_3 + q_5 + \cdots = 1$. If the walk starts at $X(0) \in B_0$, then asymptotically $X(2t) \in B_0$ with probability mass function Q_0 and $X(2t + 1) \in B_1$ with probability mass function Q_1. On the other hand, if the walk starts at $X(0) \in B_1$, then asymptotically $X(2t) \in B_1$ with probability mass function Q_1 and $X(2t + 1) \in B_0$ with probability mass function Q_0.

We invite interested readers to write codes using their favorite software to simulate the walks on Networks A, R, L, D, C; and verify the results proved in this paper.

Acknowledgment We sincerely thank the referees for their generous suggestions to improve this paper. We are indebted to the editor for polishing our earlier manuscript and to Stanley R. Huddy for redrawing our figures.

REFERENCES

[1] Chandra, A. K., Prabhakar, R., Ruzzo, W. L., Smolensky, R., Tiwari, P. (1996). The electrical resistance of a graph captures its commute and cover times, *Computational Complexity* 6(4):312–340.

[2] Chong, K. S., Cowan, R., Lars, H. (2000). The ruin problem and cover times of asymmetric random walks and Brownian motions, *Adv. in Appl. Probab.* 32(1):177–192.

[3] Doyle, P. G., Snell, J. (1984). *Random Walks and Electric Networks.* Washington, DC: Mathematical Association of America.

[4] Feller, W. (1967). *Introduction to Probability Theory and its Applications*, Vol. 1., 3rd ed. New York: Wiley.

[5] Good, I. J. (1951). Random motion in a finite Abelian group. *Math. Proc. Cambridge Philos. Soc.* 47(4):756–762.

[6] Karlin, S., Taylor, H. M. (1975). *A First Course in Stochastic Processes*, 2nd ed. San Diego, CA: Academic Press.

[7] Lovasz, L. (1993). Random walks on graphs: A survey. In: Miklos, D., Sos, V. T., Szonyi, T., eds. *Combinatorics, Paul Erdos is Eighty*, Vol. 2., Budapest, Hungary: Bolyai Society Mathematical Studies, pp. 1–46.

[8] Ross, S. M. (1996). *Stochastic Processes*, 2nd ed. New York: Wiley.

[9] Ross, S. M. (2014). *Introduction to Probability Models*, 11th ed. San Diego, CA: Academic Press.

[10] Ross, S. M. (2010). *A First Course in Probability*, 8th ed. Upper Saddle River, NJ: Pearson Prentice Hall.

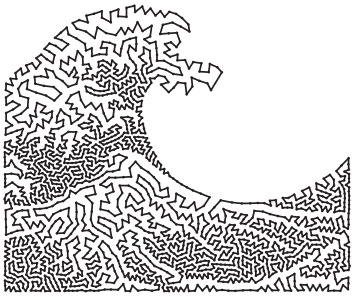
[11] Sarkar, J. (2006). Random walk on a polygon. In: *IMS Lecture Notes-Monograph Series Recent Developments in Nonparametric Inference and Probability*,Vol. 50, Beachwood, OH: Inst. Math. Statist, pp. 1–43.

[12] Vallois, P. (1996). The range of a simple random walk on $\mathbb{Z}$, *Adv. in Appl. Probab.* 28(4):1014–1033.

Summary. We study a symmetric simple random walk on a finite section of the integer lattice (with various end conditions) and on the vertices of a polygon whose nodes are labeled as $0, 1, 2, \ldots, m$. We study the probability distributions (specifically the expectation and the variance) of the time until first return to the starting position, the number of nodes visited during this time, the time until all nodes are visited, the last node to be visited, the time to return to the starting position after visiting all nodes and the number of nodes visited in the interim. These questions are answered using elementary methods readily understood by college mathematics students—methods such as symmetry, recursive relations and mathematical induction.

SARAN ISHIKA MAITI (MR Author ID: 885783) is a Senior Assistant Professor of Statistics at Visva-Bharati University, India. Her research interests are in nonparametric regression theory, multivariate statistics and stochastic processes. This paper was initiated when she was a visiting assistant professor at IUPUI. When she does not dine on statistics, she unwinds by taking long (partially random) walks on forsaken grounds.

JYOTIRMOY SARKAR (MR Author ID: 316318) is a Professor of Statistics at IUPUI. His research interests are in applied probability, combinatorics, mathematical statistics, optimization, reliability theory and statistics education. He enjoys experimenting with teaching techniques, especially on his captive audience of one—his son. In his spare time, he solves (and sometimes invents) mathematical puzzles.



Opt-emoji Hokusai, Robert Bosch; digital print, 2019. An optimal TSP tour of 2048 points that were arranged to resemble an emoji version of Hokusai's well-known wave painting. The optimal tour was obtained with the Concorde TSP Solver.

See interview on page 305.

Proof Without Words: Square Triangular Sums

CHARLES F. MARION
Yorktown Heights, NY 10598
charliemath@optonline.net

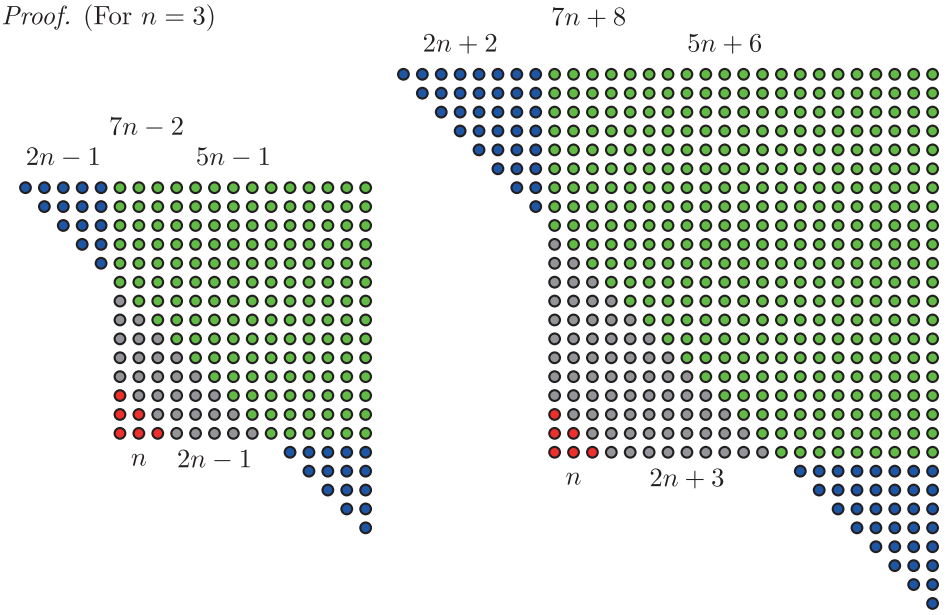
The following two patterns involving triangular numbers $T_n = 1 + \cdots + n$ suggest identities that are proved below via proofs without words:

$$1 + 15 = 4^2, \quad 3 + 78 = 9^2, \quad 6 + 190 = 14^2, \quad 10 + 351 = 19^2, \quad \ldots \text{ and}$$

$$1 + 120 = 11^2, \quad 3 + 253 = 16^2, \quad 1 + 120 = 11^2, \quad 10 + 666 = 26^2, \quad \ldots.$$

Proposition. *For* $n \in \mathbb{N}$, $T_n + T_{7n-2} = (5n-1)^2$ *and* $T_n + T_{7n+8} = (5n+6)^2$.

Proof. (For $n = 3$)



In general, suppose $a^2 + (a+1)^2 = c^2$. Then,

$$T_n + T_{(6a+4c+3)n-(a+c+1)} = \left[(4a+3c+2)n - \frac{2a+c+1}{2}\right]^2 \text{ and}$$

$$T_n + T_{(6a+4c+3)n+7a+5c+3} = \left[(4a+3c+2)n + \frac{10a+7c+5}{2}\right]^2.$$

Summary. We show that the sum of any triangular number and a term in each of two triangular number subsequences is also a perfect square.

CHARLES F. MARION (MR Author ID: 1211614; ORCID: 0000-0001-6620-3896) started teaching mathematics at Mercy College (Dobbs Ferry, NY) in 1966. He would like to acknowledge the continued support and encouragement of his long-time friends and former colleagues, Paul Hughes and Tom Guglielmo.

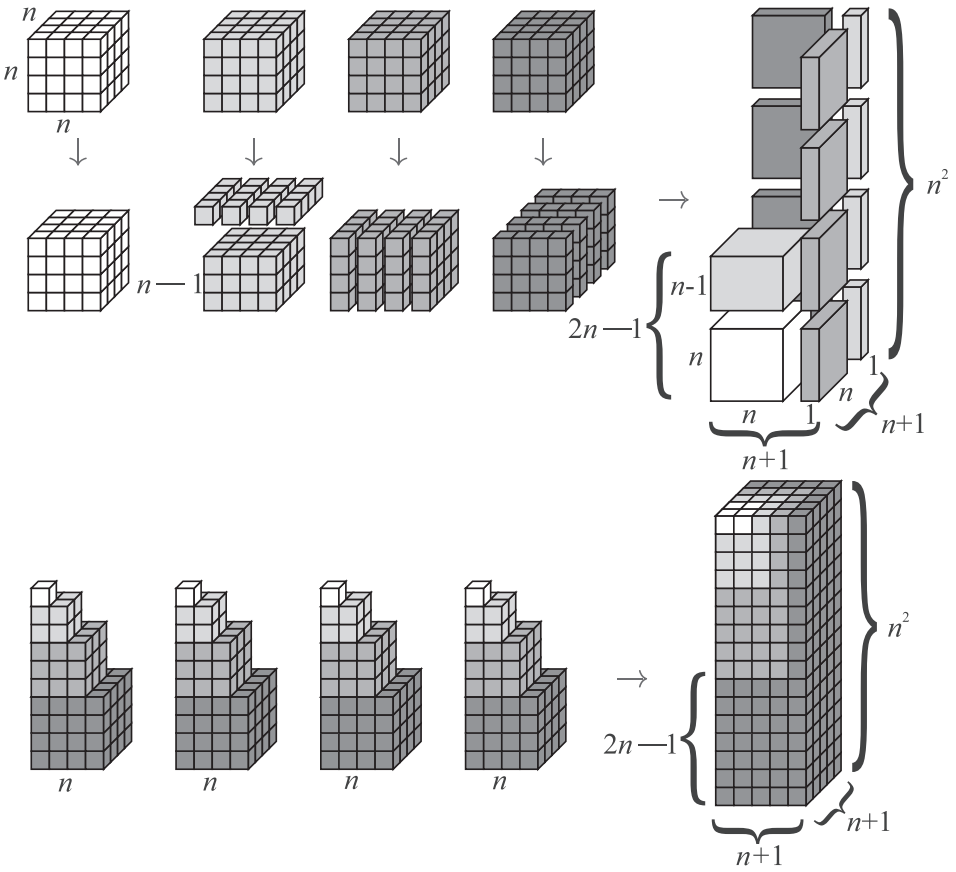
Math. Mag. **92** (2019) 269. doi:10.1080/0025570X.2019.1571374
MSC: Primary 05A19
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

Proof Without Words: Sum of Cubes

SANJA STEVANOVIĆ
DRAGAN STEVANOVIĆ
Serbian Academy of Sciences and Arts,
11001 Belgrade, Serbia
sanja_stevanovic@yahoo.com
dragance106@yahoo.com

The identity $1^3 + 2^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$ had been proved wordlessly before in several ways [1]–[3]. In this new proof, we transform cube volumes to show that

$$4(1^3 + 2^3 + \cdots + n^3) = n^2(n+1)^2.$$



Acknowledgment This work was supported by the research project ON174033 of the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Nelsen, R.B. (1993). *Proofs Without Words: Exercises in Visual Thinking.* Washington, DC: Mathematical Association of America.
[2] Nelsen, R.B. (2000). *Proofs Without Words: More Exercises in Visual Thinking.* Washington, DC: Mathematical Association of America.

Math. Mag. **92** (2019) 270–271. doi:10.1080/0025570X.2019.1571376
MSC: Primary 05A19

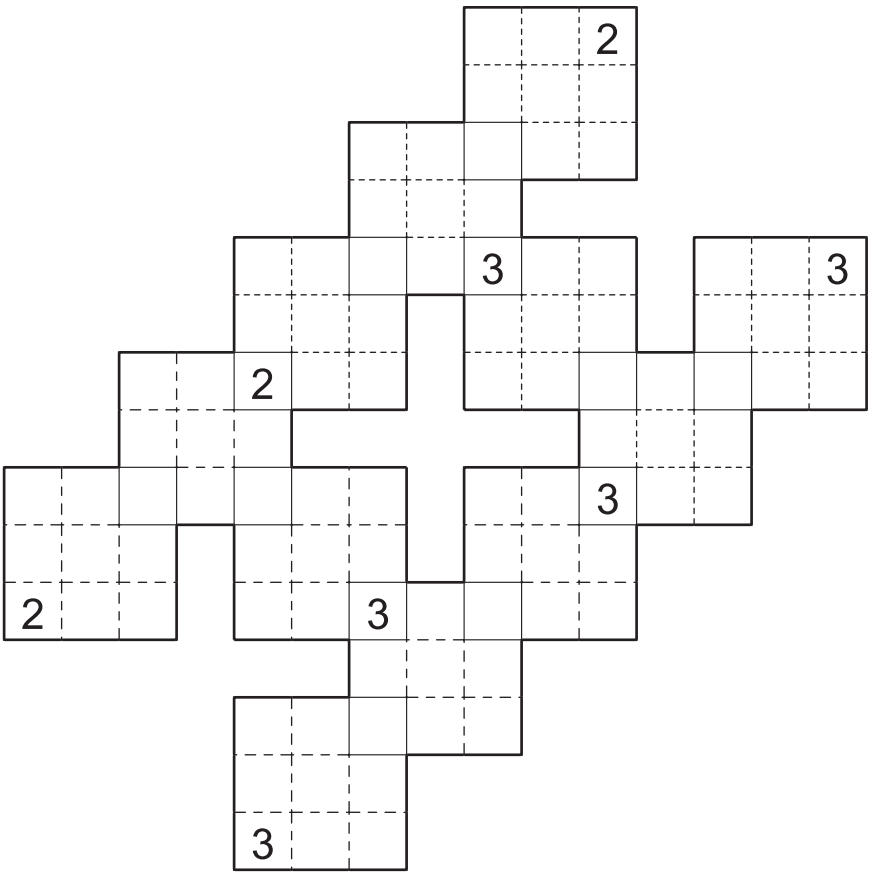
[3] Nelsen, R.B. (2016). *Proofs Without Words III: Further Exercises in Visual Thinking*. Washington, DC: Mathematical Association of America.

Summary. We transform cube volumes to show that $4(1^3 + 2^3 + \cdots + n^3) = n^2(n+1)^2$.

SANJA STEVANOVIĆ (MR Author ID: 875571) worked as an architect in industry before entering the PhD program in Architecture at University of Niš. She is an assistant research professor at Mathematical Institute of SASA with research interests in space syntax, a field applying graph theory to architecture and urbanism.

DRAGAN STEVANOVIĆ (MR Author ID: 626422) is a research professor at Mathematical Institute of SASA. His research interests are in graph theory and combinatorics.

Tribus Puzzle



How to play. Fill each of the three-by-three squares with either a 1, 2, or 3 so that each number appears exactly once in each column and row. Some cells apply to more than one square, as the squares overlap. Each of the three-by-three squares must be distinct. The solution can be found on page 287.

— David Nacin, William Paterson University, Wayne, NJ (nacind@wpunj.edu)

Triphos: A World Without Subtraction

KEELY GROSSNICKLE
Kansas State University
Manhattan, KS 66506
kgrossni@ksu.edu

BRIAN HOLLENBECK
Emporia State University
Emporia, KS 66801
bhollenb@emporia.edu

JEANA JOHNSON
Shawnee Heights Middle School
Tecumseh, KS 66542
johnsonjm@usd450.net

ZHIHAO SUN
Bank of Communications Co., Ltd.
Dalian, China
sylvester0224@hotmail.com

Imagine a world where each object consists of three attributes of varying magnitudes. The key feature of these attributes is that an equal magnitude of each is equivalent to having none. A consequence is that subtraction (and negative values) need not exist because to reduce one attribute one needs only to increase the other two. This is somewhat analogous to the interaction of colors of light, and thus we name our world *Triphos*, which is derived from the Greek words for "three" and "light," and denote the attributes R (red), G (green), and B (blue). A second feature of this imaginary world is that the fundamental shape is an equilateral triangle, instead of a square. In this work, we compare the *Triphos* world to our own, highlighting similarities and differences, and recalling some results from Egging and Johnson [**3**]. We conclude with open problems suitable for undergraduate research.

Triphosian numeration and plotting

We first need some notation to aid our exposition.

Definition 1. A *Triphosian number* can be expressed in the form ${}^{g}_{b}r$ where r, g, and b are nonnegative real numbers. If r, g, and b are nonnegative integers, then ${}^{g}_{b}r$ is said to be a *Triphosian integer*. Denote the set of all Triphosian real numbers by $\mathbb{R}_T$ and the set of all Triphosian integers by $\mathbb{Z}_T$.

The key assumption we make about Triphosian numbers is that ${}^{n}_{n}n$ and ${}^{0}_{0}0$ represent the same Triphosian number for any $n \geq 0$. As a consequence, three-dimensional space is not necessary for plotting Triphosian numbers. Instead, the most natural system is to make use of three axes on a two-dimensional plane. These three axes separate the plane into three *trients*. (The term trient was chosen since the Roman coin, *triens*, had a value of 1/3 of an *as*, while a *quadrans* had a value of 1/4 of an *as*.) The number ${}^{0}_{0}0$ is plotted at the origin where the three axes meet. An example of this coordinate system for plotting Triphosian integers is shown in Figure 1 (left).

Math. Mag. **92** (2019) 272–285. doi:10.1080/0025570X.2019.1609293
MSC: Primary 11A67
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

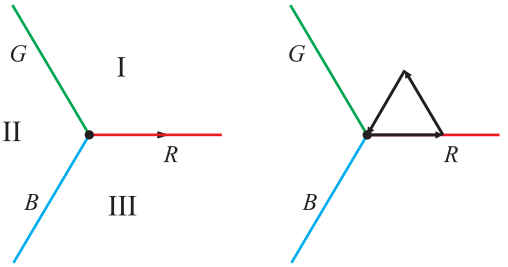


Figure 1 On the left is the origin and trients of the three-axis system. On the right is a plot of the number ${}^{3}_{3}3$.

The number ${}^{g}_{b}r$ can be represented by a point shifted r units from the origin along the R axis, then shifted g units in the direction of the G axis, and finally shifted b units in the direction of the B axis. Figure 1 (right) shows the equivalence of ${}^{0}_{0}0$ and ${}^{n}_{n}n$ and the connection between equlilateral triangles and Triphosian numbers.

Unlike the Cartesian coordinate system, points in the plane are not uniquely described. The left and middle graphs in Figure 2 show that ${}^{4}_{1}3$ and ${}^{3}_{0}2$ are equivalent. Of course, there are an infinite number of expressions that describe this same point, but only ${}^{3}_{0}2$ contains a zero for one of its components. Thus Triphosian numbers are similar to rational numbers, which motivates our next definition.

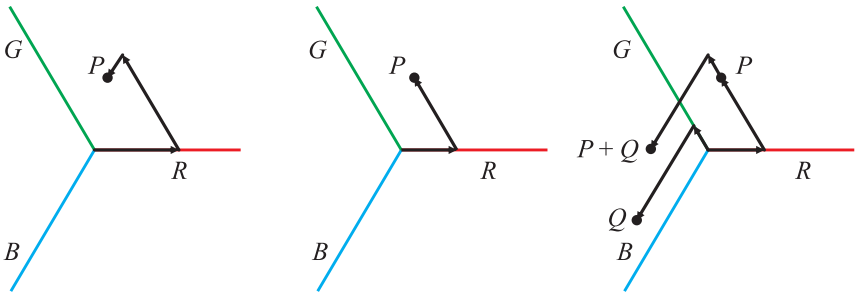


Figure 2 On the left are plots of $P = {}^{4}_{1}3$ and $P = {}^{3}_{0}2$. On the right is a plot of the sum of the points $P = {}^{3}_{0}2$ and $Q = {}^{1}_{4}0$.

Definition 2. Two Triphosian numbers are *equivalent*, ${}^{b}_{c}a \sim {}^{y}_{z}x$, if there exists an $r \in \mathbb{R}$ such that $a = x + r$, $b = y + r$, and $c = z + r$.

Because $\sim$ is an equivalence relation, as shown in [3], we choose to represent Triphosian numbers in a reduced form.

Definition 3. A Triphosian number ${}^{g}_{b}r$ is *reduced* if at least one of the components, r, g, or b, is zero.

It is clear that if $x, y > 0$, then Triphosian numbers of the form, ${}^{y}_{0}x$, lie in Trient I. Similarly, Trient II consists of numbers of the form, ${}^{x}_{y}0$, and Trient III corresponds to numbers of the form, ${}^{0}_{y}x$. If two components are zero, then the number lies on the axis corresponding to the nonzero component. Each point in the Cartesian plane can

be described by a unique reduced Triphosian number. In [3], Egging and Johnson show that a Triphosian number ${}^{g}_{b}r$ can be converted to Cartesian coordinates by matrix multiplication:

$$[r, g, b]\begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}. \tag{1}$$

Arithmetical operations

We are now ready to define arithmetical operations for our system. The analogy to light motivates the following pointwise definition for addition.

Definition 4. The *sum* of two Triphosian numbers, ${}^{b}_{c}a$ and ${}^{y}_{z}x$, is ${}^{b}_{c}a + {}^{y}_{z}x = {}^{b+y}_{c+z}a{+}x$.

The definition of addition satisfies the properties we expect. For example, the additive identity exists since by definition,

$${}^{0}_{0}0 + {}^{y}_{z}x = {}^{y}_{z}x.$$

Commutativity and associativity follow from the real numbers. The details are left to the reader. Addition should also be independent of whether the number is reduced or not. In particular, we have the following theorem.

Theorem 1. *Let $X, Y, Z \in \mathbb{R}_T$ where Y is the reduced form of X. Then $X + Z = Y + Z$.*

Proof. We prove the case where Y is in Trient I. The other cases are similar. Assume $X = {}^{b}_{c}a$, $Y = {}^{e}_{0}d$, and $Z = {}^{g}_{h}f$ where $a, b, c, d, e, f, g, h > 0$. If ${}^{b}_{c}a = {}^{e}_{0}d$, then we use the additive identity ${}^{c}_{c}c$ to see that

$${}^{b}_{c}a = {}^{e}_{0}d + {}^{c}_{c}c = {}^{e+c}_{c}d{+}c.$$

Therefore $a = d + c$ and $b = e + c$. We use the other addition properties to conclude

$$Y + Z = {}^{e}_{0}d + {}^{g}_{h}f = {}^{e}_{0}d + {}^{g}_{h}f + {}^{c}_{c}c = {}^{e+c+g}_{h+c}d{+}c{+}f = {}^{b}_{c}a + {}^{g}_{h}f = X + Z. \qquad \blacksquare$$

To prove the existence of an additive inverse of ${}^{g}_{b}r$, we "complete the triangle" by adding values to each attribute.

Theorem 2. *Let $r, g, b \geq 0$. Then an additive inverse of ${}^{g}_{b}r$ is ${}^{r+b}_{r+g}g{+}b$.*

We now search for an appropriate definition of multiplication. Unlike addition, there is no obvious parallel with the properties of light. We will distinguish between two common types of multiplication, which we call *scalar* and *Triphosian*. We first define scalar multiplication. Although we will see that this definition is redundant, it often simplifies notation.

Definition 5. Let $k \geq 0$ and $X \in \mathbb{R}_T$. If $X = {}^{g}_{b}r$ with $r, g, b \geq 0$, then we define the *scalar multiple of X* to be $kX = Xk = {}^{gk}_{bk}rk$.

For the product of two Triphosian numbers, the pointwise definition,

$${}^{b}_{c}a * {}^{y}_{z}x = {}^{by}_{cz}ax,$$

fails to have many of the familiar properties that we expect. For example, if $X = {}^{2}_{4}3$ and $Y = {}^{6}_{3}4$, then this pointwise definition yields $X * Y = {}^{12}_{12}12 = {}^{0}_{0}0$. Thus the zero product property fails to hold since neither X nor Y is the additive identity. Also, the definition of multiplication should give a consistent result for all equivalent representations of a Triphosian number, just as the multiplication of rational numbers is independent of the representation used. But if we multiply the reduced forms of X and Y we have

$$X * Y = {}^{0}_{2}1 \cdot {}^{3}_{0}1 = {}^{0}_{0}1 \neq {}^{0}_{0}0.$$

Although satisfying commutativity and associativity, we reject this definition.

Being more methodical in our search for a satisfactory definition of Triphosian multiplication, observe that, by the definition of addition and scalar multiplication, any Triphosian number, ${}^{g}_{b}r$, can be expressed using an alternative notation using the *primaries* $G = {}^{0}_{0}1$, $R = {}^{1}_{0}0$, and $B = {}^{0}_{1}0$, so that

$${}^{g}_{b}r = {}^{0}_{0}r + {}^{g}_{0}0 + {}^{0}_{b}0 = rR + gG + bB.$$

For the commutative, associative, and distributive properties to hold, we write

$$\begin{aligned} {}^{b}_{c}a \cdot {}^{y}_{z}x &= (aR + bG + cB)(xR + yG + zB) \\ &= axR{\cdot}R + ayR{\cdot}G + azR{\cdot}B + bxG{\cdot}R + byG{\cdot}G \\ &\qquad + bzG{\cdot}B + cxB{\cdot}R + cyB{\cdot}G + czB{\cdot}B. \end{aligned} \tag{2}$$

So our potential definition of multiplication has been reduced to deducing the most natural values for $R{\cdot}R$, $G{\cdot}G$, $B{\cdot}B$, $R{\cdot}G$, $G{\cdot}B$, and $B{\cdot}R$. For the sake of simplicity, we will assume the product of two primaries is also primary. Thus there are $3^6 = 729$ potential candidates for our definition. Because there is only one group of order 3 [**4**], we will eventually arrive at a single definition. However, let us examine how each property of multiplication affects our choices. For instance, we can eliminate nearly all of these 729 possibilities if we require multiplication to be well-defined, that is, independent of the representation used. For example, it is clear that ${}^{g}_{b}r \cdot {}^{0}_{0}0 = {}^{0}_{0}0$ for any Triphosian number, ${}^{g}_{b}r$. So we expect ${}^{g}_{b}r \cdot {}^{1}_{1}1 = {}^{n}_{n}n$ for some n. Using the convention $X \cdot X = X^2$ and applying (2) to the primaries, we have

$$\begin{aligned} R \cdot {}^{1}_{1}1 &= R^2 + B{\cdot}R + R{\cdot}G \\ G \cdot {}^{1}_{1}1 &= G^2 + G{\cdot}B + R{\cdot}G \\ B \cdot {}^{1}_{1}1 &= B^2 + G{\cdot}B + B{\cdot}R. \end{aligned}$$

To satisfy our assumption, each component must have an equal magnitude, implying the three terms on the right side of each equation above must be distinct. Thus,

$$\begin{aligned} R^2 &\neq B{\cdot}R \neq R{\cdot}G \\ G^2 &\neq G{\cdot}B \neq R{\cdot}G \\ B^2 &\neq G{\cdot}B \neq B{\cdot}R, \end{aligned}$$

which is equivalent to $R^2 = G{\cdot}B$, $G^2 = B{\cdot}R$, and $B^2 = R{\cdot}G$. We have reduced 729 possibilities to the six shown in Table 1.

Notice by symmetry P_2, P_3, and P_6 are equivalent. Likewise, P_4 and P_5 are equivalent. As a result, we have only three possible scenarios left for Triphosian multiplication: P_1, P_2, and P_4. But we have one more property to check, the existence of a unique

TABLE 1: Possibilities for the products of primaries

	P_1	P_2	P_3	P_4	P_5	P_6
R^2	R	R	G	B	G	B
G^2	G	B	R	R	B	G
B^2	B	G	B	G	R	R
$G\cdot B$	R	R	G	B	G	B
$B\cdot R$	G	B	R	R	B	G
$R\cdot G$	B	G	B	G	R	R

multiplicative identity. One can quickly see P_1 does not have a unique identity. Indeed, $R\cdot R = R$ implies R is a candidate for the multiplicative identity, but $R\cdot G = B \neq G$. Similarly for P_4, $R\cdot G = G$, but $R\cdot R = B \neq R$. However, R is the unique identity for multiplication defined by the products of P_2. This is true because $R\cdot R = R$, $R\cdot B = B$, and $R\cdot G = G$.

Thus the relationships of P_2 have an asymmetric quality, which we will use to define Triphosian multiplication. Using P_1 or P_4 to define multiplication values symmetry over the uniqueness of the multiplicative identity. We invite the reader to further explore these alternative definitions.

Definition 6. The *product* of the two Triphosian numbers ${}^{b}_{c}a$ and ${}^{y}_{z}x$ is

$$ {}^{b}_{c}a \cdot {}^{y}_{z}x = {}^{cz+bx+ay}_{by+az+cx}ax+bz+cy. \tag{3}$$

Example 3. Find the product of $P = {}^{3}_{0}2$ and $Q = {}^{1}_{0}2$ in reduced form. By definition, $P \cdot Q = {}^{3}_{0}2 \cdot {}^{1}_{0}2 = {}^{8}_{3}4$. Since ${}^{8}_{3}4 = {}^{3}_{3}3 + {}^{5}_{0}1$, the reduced form of the product is $S = {}^{5}_{0}1$.

Of course, we could have just as easily chosen G or B to be the multiplicative identity, but it is natural for us to place the multiplicative identity in the same position on the Triphosian plane as it would be found in the complex plane. In fact, we can immediately see the connection between Triphosian multiplication and multiplying complex numbers. It is well-known that multiplying a complex number by the imaginary unit, $i = \sqrt{-1}$, is equivalent to rotating that number 90° counterclockwise about the origin. Since the R axis of the Triphosian plane coincides with the positive real axis of the Cartesian coordinate system, then a Triphosian number of the form ${}^{0}_{0}r$ will correspond to the real number, $r > 0$. This leaves the G and B components to be "imaginary" parts, corresponding to 120° and 240° rotations, respectively. When multiplying two complex numbers, the angles associated with each number are added to find the angle corresponding to the product. Notice this fact is consistent with $G^2 = B$, along with the other primary products of P_2. Consequently, every Triphosian number, ${}^{g}_{b}r$, can be expressed as $rR + gG + bG^2$. If we replace R with 1, and G with a primitive cube root of unity, $\omega = e^{2\pi i/3} = (-1 + i\sqrt{3})/2$, then every point in the Triphosian plane corresponds to a point in the complex plane $r + g\omega + b\omega^2$, which of course coincides with (1). When $r, g, b \in \mathbb{Z}$ these numbers are known as the Eisenstein integers; see [2] for more information.

Another consequence of this definition of Triphosian multiplication is that scalar multiplication becomes redundant. This is because

$$k\left(\begin{smallmatrix} g \\ b & r \end{smallmatrix}\right) = \begin{smallmatrix} kg \\ kb & kr \end{smallmatrix} = \begin{smallmatrix} 0 \\ 0 & k \end{smallmatrix} \cdot \begin{smallmatrix} g \\ b & r \end{smallmatrix}, \text{ for all } r, g, b, k \geq 0.$$

We now show the existence of a multiplicative inverse.

Theorem 4. *Let $X = \begin{smallmatrix} g \\ b & r \end{smallmatrix}$ where $r, g, b \geq 0$. Further assume that r, g, b are not all equal. Then the multiplicative inverse of X is $X^{-1} := \frac{1}{n}Y$ where*

$$Y = \begin{smallmatrix} b \\ g & r \end{smallmatrix} \quad \text{and} \quad n = r^2 + g^2 + b^2 - rg - rb - gb.$$

Proof. We first show that the inverse is well defined by proving that $n > 0$. Reordering the terms of $2n$ shows that n is positive because

$$\begin{aligned} 2n &= r^2 - 2rg + g^2 + g^2 - 2gb + b^2 + r^2 - 2rb + b^2 \\ &= (r-g)^2 + (g-b)^2 + (r-b)^2 > 0 \end{aligned}$$

since r, g, and b are not all equal.

We now calculate

$$X \cdot \tfrac{1}{n}Y = \tfrac{1}{n}\left(\begin{smallmatrix} g \\ b & r \end{smallmatrix} \cdot \begin{smallmatrix} b \\ g & r \end{smallmatrix}\right) = \tfrac{1}{n}\left(\begin{smallmatrix} bg+gr+rb \\ gb+rg+br & r^2+g^2+b^2 \end{smallmatrix}\right) = \tfrac{1}{n}\left(\begin{smallmatrix} 0 \\ 0 & r^2+g^2+b^2-(rg+rb+gb) \end{smallmatrix}\right).$$

This last step follows since the G and B components are equal. Thus the last term can be rewritten as $\frac{1}{n}\left(\begin{smallmatrix} 0 \\ 0 & n \end{smallmatrix}\right)$. By our definition of scalar multiplication, this term is equal to the multiplicative identity $\begin{smallmatrix} 0 \\ 0 & 1 \end{smallmatrix}$. ∎

As noted in [3], the familiar properties of addition and multiplication hold for Triphosian numbers.

Theorem 5 (Egging and Johnson [3])**.** *The set of all equivalence classes in $\mathbb{R}_T$ is a field.*

Proof. The proofs in [3] appeal directly to Definitions 4 and 6. We will instead consider the following matrix representation, ρ, of Triphosian numbers, where

$$\rho\left(\begin{smallmatrix} g \\ b & r \end{smallmatrix}\right) = \begin{bmatrix} r & b & g \\ g & r & b \\ b & g & r \end{bmatrix}. \tag{4}$$

In particular, the equivalence class of $\begin{smallmatrix} g \\ b & r \end{smallmatrix}$ can be represented by the equivalence class of matrices of the form above, where two matrices are considered equivalent if all entries differ by the same constant.

Our representation satisfies our definitions for Triphosian addition and multiplication because

$$\begin{aligned} \rho\left(\begin{smallmatrix} b \\ c & a \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix} y \\ z & x \end{smallmatrix}\right) &= \begin{bmatrix} a & c & b \\ b & a & c \\ c & b & a \end{bmatrix} + \begin{bmatrix} x & z & y \\ y & x & z \\ z & y & x \end{bmatrix} = \begin{bmatrix} a+x & c+z & b+y \\ b+y & a+x & c+z \\ c+z & b+y & a+x \end{bmatrix} \\ &= \rho\left(\begin{smallmatrix} b+y \\ c+z & a+x \end{smallmatrix}\right) = \rho\left(\begin{smallmatrix} b \\ c & a \end{smallmatrix} + \begin{smallmatrix} y \\ z & x \end{smallmatrix}\right) \end{aligned}$$

and

$$\rho\left(\begin{smallmatrix} b \\ c & a \end{smallmatrix}\right) \cdot \rho\left(\begin{smallmatrix} y \\ z & x \end{smallmatrix}\right) = \begin{bmatrix} a & c & b \\ b & a & c \\ c & b & a \end{bmatrix} \cdot \begin{bmatrix} x & z & y \\ y & x & z \\ z & y & x \end{bmatrix}$$

$$= \begin{bmatrix} ax+bz+cy & az+by+cx & ay+bx+cz \\ ay+bx+cz & ax+bz+cy & az+by+cx \\ az+by+cx & ay+bx+cz & ax+bz+cy \end{bmatrix} = \rho\left(\begin{smallmatrix} b \\ c \end{smallmatrix} a \cdot \begin{smallmatrix} y \\ z \end{smallmatrix} x\right).$$

This gives the following matrix representations for familiar Triphosian numbers:

$$\rho\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 0\right) = \mathbf{0} := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \rho\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 1\right) = \mathbf{I} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\rho\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix} 0\right) = \mathbf{G} := \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \rho\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix} 0\right) = \mathbf{B} := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

It is not hard to see that $\mathbf{B} = \mathbf{G}^2$, and therefore $\rho\left(\begin{smallmatrix} g \\ b \end{smallmatrix} r\right) = r\mathbf{I} + g\mathbf{G} + b\mathbf{B}$. The advantage of the matrix representation is we can use known results about matrix addition and multiplication to simplify our proof. For instance, it is well known that matrix addition is commutative and associative with $\mathbf{0}$ as the additive identity. We have already seen in the previous section that the additive inverse exists. It is also known that matrix multiplication is associative and distributive, but not necessarily commutative. However, due to the symmetry of our particular matrix representation, commutativity holds as can be checked by the reader. Since $\mathbf{I}$ is clearly the multiplicative identity, we only need to show the existence of the multiplicative inverse for all nonzero elements. A matrix is invertible if and only if its determinant is nonzero. A simple calculation shows the determinant of the matrix of (4) is

$$r^3 + g^3 + b^3 - 3rgb = (r+g+b)(r^2+g^2+b^2-rg-rb-gb).$$

Therefore, the determinant of a matrix not equal to $\mathbf{0}$ will be zero if and only if $r^2 + g^2 + b^2 - rg - rb - gb = 0$. We have already seen in Theorem 4 that this equality holds only when $r = g = b$, which corresponds to the equivalence class of $\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 0$. Thus the theorem is proved. ■

We now confirm the zero product property holds for this definition of multiplication.

Theorem 6 (Zero product property)**.** *$X \cdot Y = \begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 0$ if and only if for some $n \geq 0$, $X = \begin{smallmatrix} n \\ n \end{smallmatrix} n$ or $Y = \begin{smallmatrix} n \\ n \end{smallmatrix} n$ or both.*

Proof. Without loss of generality, it is enough to assume $X = \begin{smallmatrix} n \\ n \end{smallmatrix} n$ and let $Y = \begin{smallmatrix} b \\ c \end{smallmatrix} a$ where $a, b, c \geq 0$. From Definition 6 we have $X \cdot Y = \begin{smallmatrix} cn+an+bn \\ bn+an+cn \end{smallmatrix} an+bn+cn = \begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 0$, since all three components are equal.

To prove the converse, we assume $X \cdot Y = \begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 0$. If $X = \begin{smallmatrix} n \\ n \end{smallmatrix} n$, then we are done. If not, then X has a multiplicative inverse. Multiplying both sides of $X \cdot Y = \begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 0$ by this inverse yields $\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 1 \cdot Y = \begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 0$. Since $\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} 1$ is the multiplicative identity, the theorem is proved. ■

Finally, we would like to check that our definition of multiplication yields a product that is independent of whether the number is in reduced form or not. Similar to our analysis of addition, we have the following theorem.

Theorem 7. *Let $X, Y, Z \in \mathbb{R}_T$ where Y is the reduced form of X. Then $X \cdot Z = Y \cdot Z$.*

Proof. If Y is the reduced form of X, then there exists an $n > 0$ such that $Y + {}^{n}_{n}{}^{n} = X$. By the distributive and zero-product properties, we conclude

$$X \cdot Z = (Y + {}^{n}_{n}{}^{n}) \cdot Z = Y \cdot Z + {}^{n}_{n}{}^{n} \cdot Z = Y \cdot Z + {}^{0}_{0}{}^{0} = Y \cdot Z.$$

■

Armed with a satisfactory definition of multiplication, it is not hard to solve linear equations.

Example 8. To solve the equation $AX + B = C$ for X where $A = {}^{2}_{0}{}^{3}$, $B = {}^{1}_{2}{}^{0}$, and $C = {}^{0}_{8}{}^{13}$, we add the additive inverse of B to both sides of the equation. Theorem 2 gives this number to be ${}^{2}_{1}{}^{3}$. We also need the multiplicative inverse of A, which is $\frac{1}{7}({}^{0}_{2}{}^{3})$ by Theorem 4. Hence, $AX + B = C$ is equivalent to

$$X = \tfrac{1}{7}({}^{0}_{2}{}^{3}) \cdot ({}^{0}_{8}{}^{13} + {}^{2}_{1}{}^{3}) = \tfrac{1}{7}({}^{0}_{2}{}^{3}) \cdot {}^{2}_{9}{}^{16} = \tfrac{1}{7}({}^{0}_{2}{}^{3}) \cdot {}^{0}_{7}{}^{14} = {}^{0}_{2}{}^{3} \cdot {}^{0}_{1}{}^{2} = {}^{2}_{7}{}^{6} = {}^{0}_{5}{}^{4}.$$

Triphosian primes

A natural progression from multiplication is factoring. To be able to factor, one must first understand Triphosian primes. How should one define a prime number in this system? Once again, we look to the complex numbers for inspiration. In particular, the *Gaussian integers*, $\mathbb{Z}[i]$, are of the form $a + bi$ where a and b are integers [5]. Even though we know 5 as a real number is prime, 5 as a Gaussian integer is not, since it is the product of $2 + i$ and $2 - i$. On the other hand, $3i$ is still characterized as prime, since i is considered to be a *unit*, rather than a separate factor. There are four units in the Gaussian integers [5]. Thus, $3i$ is known as an *associate* of 3; -3 and $-3i$ are also associates of 3. These ideas motivate the following definitions.

Definition 7. A Triphosian integer is a *unit* if it has a multiplicative inverse in $\mathbb{Z}_T$.

Definition 8. Two Triphosian integers X and Y are *associates* if there is a unit, $U \in \mathbb{Z}_T$, such that $X = U \cdot Y$.

Definition 9. A Triphosian integer is *prime* if it has no non-unit factors in $\mathbb{Z}_T$.

Let us focus on primes of the form, ${}^{0}_{0}{}^{p}$. Since ${}^{0}_{0}{}^{a} \cdot {}^{0}_{0}{}^{b} = {}^{0}_{0}{}^{ab}$, we only need to investigate prime values of p. To aid our discussion, we need to calculate the distance from the origin to any point in $\mathbb{R}_T$ expressed in reduced form. We define this distance by applying the law of cosines to a 60° angle.

Definition 10. Let $X \in \mathbb{R}_T$ be of one of the reduced forms ${}^{b}_{0}{}^{a}$, ${}^{0}_{b}{}^{a}$, or ${}^{a}_{b}{}^{0}$, $a, b \geq 0$. Then the *norm* of X is $\|X\| = \sqrt{a^2 + b^2 - ab}$.

To analyze Triphosian primes, we first need to find the units of $\mathbb{Z}_T$.

Theorem 9. *The six units for the Triphosian integers are*

$${}^{0}_{0}{}^{1}, \quad {}^{1}_{1}{}^{0}, \quad {}^{1}_{0}{}^{0}, \quad {}^{0}_{1}{}^{1}, \quad {}^{0}_{1}{}^{0}, \quad {}^{1}_{0}{}^{1}.$$

Proof. It is easy to see that each pair above are multiplicative inverses of each other. Hence by definition, all six must be units. To show that no other Triphosian integer can be a unit, consider $X \in \mathbb{Z}_T$ in one of the reduced forms of Definition 10 with

$a > 1$ or $b > 1$ or both. Let X^{-1} equal the multiplicative inverse of X. Then $\|X\| = \sqrt{a^2 + b^2 - ab} \neq 0$ and by Theorem 4,

$$\|X^{-1}\| = \frac{\sqrt{a^2 + b^2 - ab}}{a^2 + b^2 - ab} = \frac{1}{\sqrt{a^2 + b^2 - ab}} < 1,$$

where the last inequality is true due to the restrictions on a and b. Since no Triphosian integer can be less than 1 unit from the origin, $X^{-1} \notin \mathbb{Z}_T$, and thus X is not a unit. ■

One consequence of Theorem 9 is if ${}^{0}_{0}p$ is prime, then so are its associates:

$${}^{p}_{0}p, \quad {}^{0}_{p}p, \quad {}^{p}_{0}0, \quad {}^{0}_{p}0, \quad \text{and} \quad {}^{p}_{p}0.$$

Notice that all six of these primes correspond to points plotted on the Cartesian plane which lie on a single circle of radius, p, centered at the origin. What can we say about p? Since we have already seen that all elements of $\mathbb{Z}_T$ are an alternative way to represent the Eisenstein integers, then we can use known results about Eisenstein (or Eisenstein-Jacobi) primes. In particular, we have the following from [2].

Remark. The Triphosian integer ${}^{0}_{0}p$ is prime if and only if $p = 2$ or p is prime and of the form $p = 6n + 5$ for some integer $n \geq 0$.

Thus ${}^{0}_{0}5$ and ${}^{0}_{0}11$ are prime, but ${}^{0}_{0}3$ is not prime since ${}^{0}_{0}3 = {}^{1}_{2}0 \cdot {}^{2}_{1}0$. It is possible for ${}^{0}_{0}3$ to be factored using other reduced Triphosian numbers, but the factors are associates of ${}^{1}_{2}0$ and ${}^{2}_{1}0$. On the other hand, notice

$${}^{0}_{0}7 = {}^{3}_{0}1 \cdot {}^{0}_{3}1 = {}^{3}_{0}2 \cdot {}^{0}_{3}2.$$

So ${}^{0}_{0}7$ does not have a unique factorization over the set of Triphosian primes since ${}^{3}_{0}2$ is not an associate of ${}^{3}_{0}1$ or ${}^{0}_{3}1$. (As noted in [2], a unique factorization is possible for all nonzero Triphosian numbers if factors are restricted to powers of ${}^{1}_{1}0$, ${}^{1}_{0}0$, ${}^{b}_{0}a$, and ${}^{0}_{b}a$ where $b \leq \frac{a}{2}$.) The key to finding these factorizations is to observe

$${}^{0}_{0}p = {}^{b}_{0}a \cdot {}^{0}_{b}a, \quad \text{where} \quad p = \|{}^{b}_{0}a\| \cdot \|{}^{0}_{b}a\| = a^2 + b^2 - ab, \tag{5}$$

and find a pair of values, a and b, which generate $p = a^2 + b^2 - ab$. For the case $p = 7$, there are two different pairs of values, a and b, that will generate $a^2 + b^2 - ab = 7$. In fact, ${}^{0}_{0}7$ has 12 prime factors, as shown in Figure 3 (left). Notice all 12 factors lie on the same circle, centered at the origin, with a radius of $\sqrt{7}$.

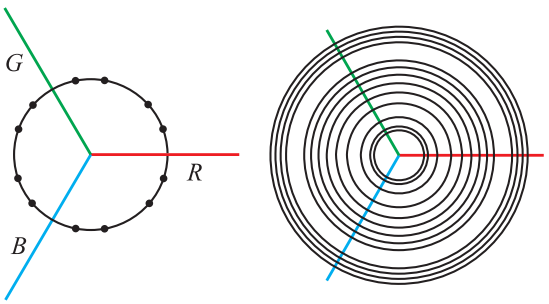


Figure 3 On the left is a plot of the 12 factors of ${}^{0}_{0}7$. On the right is a plot of prime circles of Triphosian integers.

This observation leads to a simple method for factoring Triphosian numbers of the form ${}^{0}_{0}p$, where p is prime. Draw a circle of radius $\sqrt{p}$, centered at the origin. Any Triphosian integer that lies on that circle will be a prime factor. If no integers lie on the circle, then ${}^{0}_{0}p$ is prime itself. This is because all primes, $p > 2$, are either congruent to 1 mod 6 or congruent to 5 mod 6. As we have seen, those values of p congruent to 5 mod 6 correspond to the Triphosian primes, ${}^{0}_{0}p$. Those values of p congruent to 1 mod 6 correspond to the factorization in (5). This leads to the following remark.

Remark. If one Triphosian integer lying on a circle centered at the origin is prime, then every other Triphosian integer lying on that circle is also prime.

Figure 3 (right) shows these circles for all radii less than 9.

The metric used for the norm is certainly helpful for understanding Triphosian primes. However, Egging and Johnson [3] provide another logical candidate.

Definition 11 (Egging and Johnson [3])**.** The *hexa-metric* function $D : \mathbb{R}_T \times \mathbb{R}_T \mapsto \mathbb{R}$ is defined by

$$\begin{aligned} D({}^{b}_{c}a, {}^{y}_{z}x) = \min\{&|(a-c)-(x-z)| + |(b-c)-(y-z)|, \\ &|(a-b)-(x-y)| + |(c-b)-(z-y)|, \\ &|(b-a)-(y-x)| + |(c-a)-(z-x)|\}. \end{aligned}$$

This definition calculates the shortest path between two points, using only paths parallel to the three axes. Egging and Johnson [3] show that the hexa-metric is indeed a metric, along with some other results.

Square-base system

As noted earlier, the mathematics of *Triphos* can develop naturally from its basic premises with as little influence from our own mathematics as possible. We now want to discuss what might develop from the second premise that an equilateral triangle is the fundamental shape of *Triphos*, rather than a square. We choose an equilateral triangle because the role it plays when finding an additive inverse of a Triphosian number, X. Recall a 60° angle was also key when defining the norm of X. We immediately have a new definition of area.

Definition 12. The *area* of an equilateral triangle with a side length of 1 is equal to 1.

This definition complicates finding the area of a rectangle, but calculating areas of equilateral triangles and parallelograms with 60° angles is simple. For example, the areas of the triangles in Figure 4 form a sequence of perfect squares: 1, 4, 9, 16,.... We will use this sequence to motivate our choice of representation of the magnitude of each component of a Triphosian integer. We have no reason to assume a base-10 representation, nor any other base, would be used on *Triphos*. So we assume the Triphosians employ a place value system based on perfect squares, rather than powers of 10. We now investigate the consequences of such a choice.

Observe that a representation in terms of perfect squares is not unique. For example,

$$14 = 9 + 4 + 1 = 4 + 4 + 4 + 1 + 1 = 9 + 1 + 1 + 1 + 1 + 1.$$

To parallel the familiar base-10 system more closely, let us rewrite this as

$$14 = 1 \cdot 3^2 + 1 \cdot 2^2 + 1 \cdot 1^1 = 3 \cdot 2^2 + 2 \cdot 1^2 = 1 \cdot 3^2 + 5 \cdot 1^2.$$

To simplify notation further, we adopt a place value system, using periods to distinguish it from the base-10 representation.

Figure 4 Equilateral triangles of integer side lengths.

Definition 13. The *square-base* notation $...a_5.a_4.a_3.a_2.a_1$ represents the value

$$...a_5 \cdot 5^2 + a_4 \cdot 4^2 + a_3 \cdot 3^2 + a_2 \cdot 2^2 + a_1 \cdot 1^2.$$

We can now express 14 in several concise ways:

$$14 = 1.1.1 = 3.2 = 1.0.5.$$

Notice we may require $0 \le a_j \le 3$ for each $j \ge 1$. This is because $4j^2 \ge (j+1)^2$ for $j \ge 1$. In fact, $2j^2 > (j+1)^2$ for $j \ge 3$. Thus the digits 0 or 1 would suffice for any digit, a_j, where $j \ge 3$. But for the sake of simplicity, we will adopt the convention that any digit must be between zero and three. Thus 1.1.1 or 3.2 would be acceptable representations for 14, but not 1.0.5.

One benefit of the square-base system is that any integer can be represented by at most four non-zero digits. This is due to Lagrange's four-square theorem in which every positive integer can be written as the sum of four squares [1]. A disadvantage is there may be long strings of zeros between each non-zero digit. For example, 51 can be expressed as 1.0.0.0.0.0.2. To shorten this string of zeros, we will exponentiate zero to indicate the total number of zeros in the string. For example, $51 = 1.0^5.2$. In general, we use the notation $a.0^n.b = a.0.0.\cdots.0.b$ where the string between a and b consists of n zeros.

One benefit of the base-10 system is the algorithms for addition and multiplication are relatively simple. We next investigate these algorithms for the square-base system.

Addition. We first note that square-base addition is similar to base-10 addition since we can add digits of the same place value. We may write

$$\begin{array}{r} 1.1.1 \\ +\ 1.0.1.0 \\ \hline 1.1.2.1 \end{array}$$

in the square-base system. One difference, however, is the process of "carrying." In base-10, we carry if the sum of the digits for one place is greater than 9. By convention, we have decided to carry in the square-base system if the sum is greater than 3. But we cannot carry to the next column, as we do in base-10. It is helpful to be aware of the basic addition facts that involve carrying in the square-base system:

$$1+3 = 1.0, \quad 2+2 = 1.0, \quad 2+3 = 1.1, \quad \text{and} \quad 3+3 = 1.2.$$

Numerical experimentation indicates we should carry from column j to column $2j$, where columns are counted from the right. For example,

$$\begin{array}{r} {\scriptstyle 1} \\ 23 \\ +\ 19 \\ \hline 42 \end{array} \quad \text{becomes} \quad \begin{array}{r} {\scriptstyle 1\ \ 1} \\ 1.0.1.3 \\ +\ 1.2.2 \\ \hline 2.1.0.1 \end{array}$$

where we carried from the first column to the second column, and from the second column to the fourth column. This holds in general, because if we let $x = \sum_{k=1}^{n} a_k k^2$ and $y = \sum_{k=1}^{n} b_k k^2$ where $4 \le a_j + b_j \le 6$ for some $1 \le j \le n$ and $0 \le a_k + b_k \le 3$ for all $k \ne j$, then

$$x + y = (a_1 + b_1)1^2 + (a_2 + b_2)2^2 + \cdots + (a_n + b_n)n^2.$$

Since $a_j + b_j > 3$, there is an $x \in \{0, 1, 2\}$ such that $a_j + b_j = 4 + x$. So the jth term becomes $(4 + x)j^2 = 1(2j)^2 + xj^2$. Thus the coefficient of the jth term becomes x and the coefficient of the $2j$th term is 1. Hence the extra digit must be carried to the column twice the number of the original place value.

Multiplication. The algorithm for square-base multiplication also has some similarities and differences. We first note that the square-base multiplication table shown in Table 2 is much smaller than the more familiar 9×9 grid needed for base-10 multiplication. Notice that we have adopted the convention of $3 \times 3 = 2.1$ instead of $3 \times 3 = 1.0.0$. Either form is correct, although the former is more concise. Nevertheless, multiplication results in long expressions as shown in the following examples.

TABLE 2: The square-base multiplication table

×	1	2	3
1	1	2	3
2	2	1.0	1.2
3	3	1.2	2.1

Example 10. We write the products $12 \times 3 = 36$, $27 \times 3 = 81$, and $12 \times 12 = 144$ in square-base notation.

$$\begin{array}{r} 3.0 \\ \times\ 3 \\ \hline 2.0.1.0 \end{array} \qquad \begin{array}{r} 3.0.0 \\ \times\ 3 \\ \hline 2.0.0.1.0.0 \end{array} \qquad \begin{array}{r} 3.0 \\ \times\ 3.0 \\ \hline 2.0.0.0.1.0.0.0 \end{array}$$

It is evident that the "2" and the "1" came from the multiplication table. From these examples we conjecture that if the "1" is in the jth column, then "2" is placed in the $2j$th column, where j is the product of the number of digits in each number. We formalize these observations in the following theorem.

Theorem 11. *Let two numbers x, y with m digits and n digits, respectively, have a square-base representation of $x = a.0^{m-1}$ and $y = b.0^{n-1}$ where $1 \le a, b \le 3$. Then the square-base representation of xy has nonzero digits at only the digits corresponding to columns mn and $2mn$.*

Proof. Case 1: ($a = 3$ and $b = 3$). By definition, $x = 3m^2$ and $y = 3n^2$. Therefore

$$xy = 9m^2n^2 = 2(2mn)^2 + 1(mn)^2 = 2.0^{mn-1}.1.0^{mn-1}.$$

Case 2: ($a = 3$ and $b = 2$). For this case, we let $x = 3m^2$ and $y = 2n^2$. Then

$$xy = 6m^2n^2 = 1(2mn)^2 + 2(mn)^2.$$

Case 3: ($a = 2$ and $b = 2$). Let $x = 2m^2$ and $y = 2n^2$. Then

$$xy = 4m^2n^2 = 1(2mn)^2.$$

Case 4: ($a = 1$ or $b = 1$ or both). Without loss of generality, assume $a = 1$, and $b \in \{1, 2, 3\}$. Let $x = m^2$ and $y = bn^2$. Then $xy = bm^2n^2 = b(mn)^2$.

Notice in all four cases, the only nonzero terms corresponded to $(2mn)^2$ or $(mn)^2$ or both, proving the theorem. ■

So far we have only analyzed multiplication of two numbers, each consisting of one nonzero digit. As is true with base-10 multiplication, this is sufficient to multiply numbers with more nonzero digits. As an example, let us look at the general case of multiplying a three-digit number by a two-digit number.

$$\begin{aligned} a.b.c \times d.e &= (a \cdot 3^2 + b \cdot 2^2 + c \cdot 1^2)(d \cdot 2^2 + e \cdot 1^2) \\ &= (a \cdot 3^2)(d \cdot 2^2) + (a \cdot 3^2)(e \cdot 1^2) + (b \cdot 2^2)(d \cdot 2^2) \\ &\qquad + (b \cdot 2^2)(e \cdot 1^2) + (c \cdot 1^2)(d \cdot 2^2) + (c \cdot 1^2)(e \cdot 1^2) \\ &= a.0.0 \times d.0 + a.0.0 \times e + b.0 \times d.0 + b.0 \times e + c \times d.0 + c \times e \end{aligned}$$

Thus we have reduced this problem to six simpler multiplication problems.

Example 12. We calculate $31 \times 11 = 341$ using the square-base algorithm. In square-base notation, $31 = 2.3.1$ and $11 = 2.3$. So we first calculate the products needed to find $2.3.1 \times 3$.

2.0.0	3.0	1
$\times$ 3	$\times$ 3	$\times$ 3
1.0.0.2.0.0	2.0.1.0	3

Next we calculate the products needed to find $2.3.1 \times 2.0$.

2.0.0	3.0	1
$\times$ 2.0	$\times$ 2.0	$\times$ 2.0
1.0.0.0.0.0.0.0.0.0.0.0	1.0.0.0.2.0.0.0	2.0

Adding these six results using the square-base algorithm yields 1.0.0.0.2.0.1.0.0.2.3.3, which is 341 in base-10, as expected.

Could a better representation scheme than the square-base system develop on *Triphos*? One possible compromise could be the notation would eventually move to base-4, which would retain the smaller multiplication table. In this case, our multiplication problem becomes $133_4 \times 23_4 = 11111_4$.

Future work

We invite the reader to consider other avenues of research in the *Triphos* system. We list a few potential areas to investigate below.

Trigonometry The more natural measure of a "right" angle on *Triphos* would be 60°, instead of 90°. Thus a new version of the Pythagorean theorem for triangles with at least one 60° angle could be formulated. Also, new trigonometric functions could be defined based on this new definition of a "right" triangle.

Graphing This could be investigated using more than one approach. For example, one notices that a plot of the points of the form ${}^{x}_{0}1$, $x > 0$, creates a line parallel to the G axis. How would one characterize a plot of the form ${}^{f(x)}_{0}a$, where a is a constant, and f is a function of a real number x? Alternatively, one could analyze mappings such as $Y = X^2$, where $X, Y \in \mathbb{R}_T$. This would likely require an approach similar to what is done when illustrating a transformation involving complex variables by using two separate planes.

Geometry Standard area formulas are based on measuring in square units. How would these formulas change if the basic unit was an equilateral triangle of side length equal to one? Also, what results would change using a concept of distance based on the hexa-metric?

Place-value notation Are there any other insights to be gained from the square-base system? Would place-values based on the triangular numbers, 1, 3, 6, 10, ... be an improvement?

Fractals Complex numbers are used to generate fractals such as the Julia set. What fractals could be generated using Triphosian numbers?

We are confident interested readers could find other topics to explore besides these. We hope it will be a fruitful experience that will allow the reader to better understand and appreciate traditional mathematics in the process.

Acknowledgments The authors would like to thank the referees and editors for several helpful suggestions that greatly improved the quality of this paper.

REFERENCES

[1] Burton, D. M. (1998). *Elementary Number Theory*, 4th ed. New York: McGraw-Hill, p. 250.
[2] Conway, J. H., Guy, R. K. (1996). *The Book of Numbers*. New York: Springer-Verlag, pp. 220–223.
[3] Egging, P., Johnson, E. (2015). Triphos: An alternative coordinate system. *The Pentagon*. 75: 16–45.
[4] Fraleigh, J. B. (1994). *A First Course in Abstract Algebra*, 5th ed. Reading, MA: Addison-Wesley, p. 54.
[5] Gethner, E., Wagon, S., Wick, B. (1998). A stroll through the Gaussian primes. *Amer. Math. Monthly*. 105: 327–337.

Summary. We explore the *Triphos* system which results from two premises. First, we assume the subtraction operation is unnecessary because the concept of number is based on a "reality" of three attributes. Cancellation occurs when all three magnitudes of these attributes are equal. We also assume the most fundamental shape is an equilateral triangle instead of a square. We construct new definitions and prove theorems that are more natural in the *Triphos* system, and discuss some potential avenues for future research.

KEELY GROSSNICKLE (MR Author ID: 1262714) is a graduate of Emporia State University, earning Bachelors of Science degrees in both Mathematics and Economics in 2013. She will complete her Ph.D. in Mathematics at Kansas State University in 2019. She enjoys Zumba and traveling, having visited 24 countries.

BRIAN HOLLENBECK (MR Author ID: 667079) is a professor of mathematics and chair of the Department of Mathematics and Economics at Emporia State University. He received his Ph.D. from the University of Missouri in 2001. He enjoys mathematical magic and 3D printing, as well as visiting national parks with his family.

JEANA JOHNSON (MR Author ID: 1309158) is a graduate of Emporia State University, where she earned a Bachelor of Science degree in Secondary Mathematics Education. She is now a middle school math teacher at Shawnee Heights Middle School, where she hopes to inspire her students with her love of mathematics and instill in them strong mathematical thinking skills.

ZHIHAO SUN (MR Author ID: 1309159) is a graduate of Emporia State University, earning a Bachelor of Science degree in Mathematics in 2014. Now he is an Asset Manager at a local bank in China.

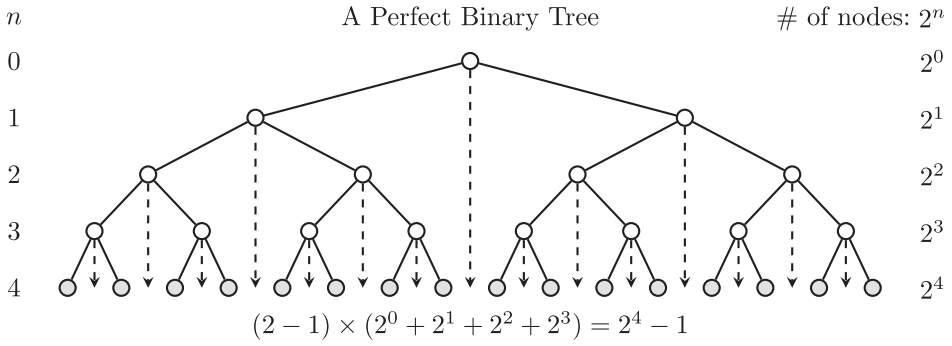
Proof Without Words: Sums of Consecutive Powers of k Via a Perfect k-ary Tree

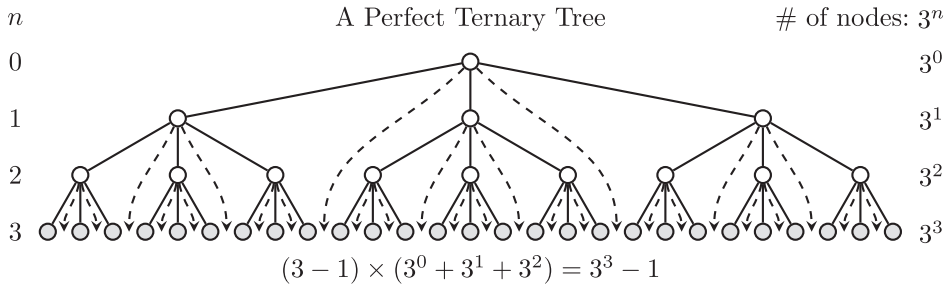
MINGJANG CHEN
Center for General Education
National Chiao Tung University
Hsinchu, Taiwan
mjchen@mail.nctu.edu.tw

Theorem. *For integers $k > 1$ and $n \geq 1$,*

$$(k-1) \times (k^0 + k^1 + \cdots + k^{n-1}) = k^n - 1.$$

This identity is proved without words using a perfect k-ary tree, in which every node other than the leaves has exactly k children and all the leaves are at the same level. Counting each of the number of internal nodes of a perfect k-ary tree $k-1$ times is equal to the number of leaves of the tree minus 1. This is proved without words below for a perfect binary tree ($k = 2$ and $n = 4$) and for a ternary tree ($k = 3$ and $n = 3$). For another proof of this identity, see [1].





REFERENCES

[1] Chen, M. (2004). Sums of consecutive powers of n via self-similarity. *Math. Mag.* 77(5): 373.

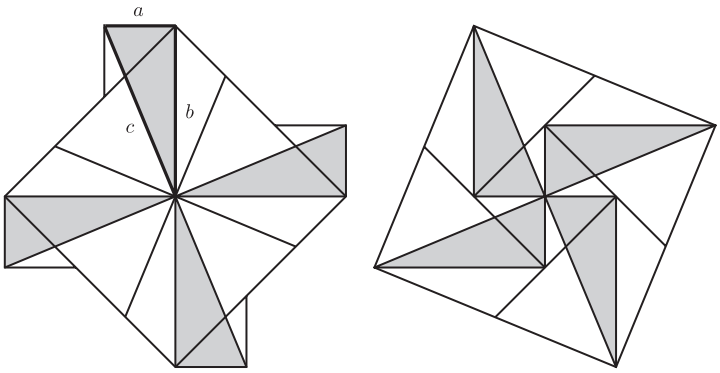
Summary. By counting the nodes of a perfect k-ary tree, an identity involving the sums of consecutive powers of k is proved.

MINGJANG CHEN (MR Author ID: 681503) received his master's and Ph.D. degrees from National Chiao Tung University, Taiwan, in 1985 and 1999, respectively. He is a full professor in the Center for General Education. His current research interests include learning with cognitive considerations, math, and art.

Math. Mag. **92** (2019) 286. doi:10.1080/0025570X.2019.1637663
MSC: Primary 05A19

Proof Without Words: The Pythagorean Theorem

JOHN MOLOKACH
Johnston Community College
Smithfield, NC 27577
johnmolokach@gmail.com



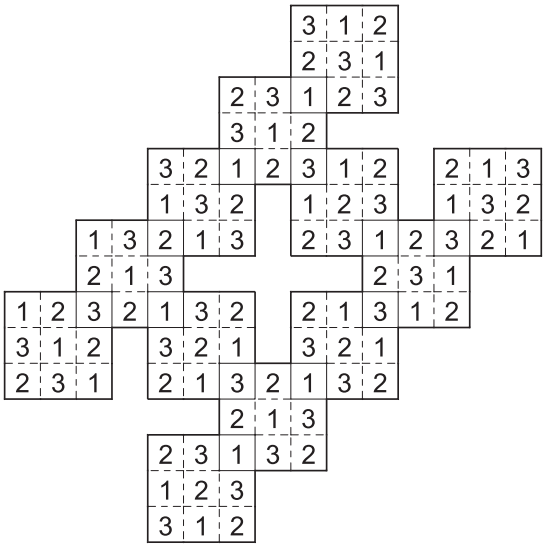
$$2a^2 + 2b^2 = 2c^2 \quad \Rightarrow \quad a^2 + b^2 = c^2$$

Summary. We present a wordless proof of the Pythagorean theorem by rearranging pieces to form a single square.

JOHN MOLOKACH (MR Author ID: 1016584) is a high school mathematics teacher at Smithfield-Selma High School in Smithfield, NC and is also an adjunct instructor at Johnston Community College, also in Smithfield, NC. He is an active AP Calculus Exam Reader and has published several wordless proofs and short articles in MAA publications, as well as acting as a reviewer for Mathematics Magazine. John also has authored several number sequences at OEIS. He holds a Master's Degree in Mathematics from East Carolina University in Greenville, NC, where his area of research was in analytic number theory. John also has an undergraduate degree in Secondary Mathematics Education from North Carolina State University.

Math. Mag. **92** (2019) 287. doi:10.1080/0025570X.2019.1586232
MSC: Primary 51M25

TRIBUS Solution



A Result on Polynomials Derived Via Graph Theory

ROBERT S. COULTER
University of Delaware
Newark, DE 19716
coulter@udel.edu

STEFAAN DE WINTER
Michigan Technological University
Houghton, MI 49931
sgdewint@mtu.edu

ALEX KODESS
Farmingdale State College
Farmingdale, New York 11735
alex.kodess@farmingdale.edu

FELIX LAZEBNIK
University of Delaware
Newark, DE 19716
fellaz@udel.edu

Graph theory is a comparatively young mathematical discipline. It is often hard to construct graphs that satisfy certain properties purely combinatorially, i.e., by taking a set of vertices and saying which vertex is connected to which. Often such areas of classical mathematics as number theory, geometry, or algebra are used for this, and the methods from the related areas are used to prove the properties of the obtained graphs. The examples are numerous, and many of them can be found in books and comprehensive survey articles. See, for example, Alon [**1**]; Babai and Frankl [**3**]; Biggs [**5**]; Füredi and Simonovits [**11**]; Brouwer and Haemers [**7**], and Alon and Spencer [**2**]. Here we wish to mention just a few such applications. The probabilistic method was used to prove the existence of certain graphs in Ramsey theory, and explicit constructions for these graphs are still unknown (see [**2**]). Constructions and analysis of Ramanujan graphs are often based on algebra and number theory. Methods of linear algebra are fundamental for studies of expanders and graphs with high degree of symmetry (see [**3**] and [**7**]). Lovász's proof [**19**] of a conjecture on the chromatic number of Kneser graphs made use of algebraic topology.

Can the direction be reversed, i.e., can graph theory be used to obtain results in some classical areas of mathematics? Sometimes it can, but the number of such instances is much smaller. See, for example, Swan's proof of the Amitsur-Levitzki theorem [**23**], or a counterexample to Borsuk's conjecture by Kahn and Kalai [**13**] and related work by Bondarenko [**6**]. Extremal graph theory was used in probability by Katona [**14**], and in geometry and potential theory by Turán [**24**], and Erdős, Meir, Sos, and Turán [**10**]. For some applications of graph theory to linear algebra, see Doob [**9**]. A number of applications of graph theory to pure mathematics are mentioned in Lovász, Pyber, Welsh and Ziegler [**20**]. The story we wish to share is about one such example. It was discovered entirely not by design.

In order to describe our problem, we need a few preliminaries. Let p be a prime number, and $\mathbb{Z}_p = \{0, 1, 2, \ldots, p-1\}$ be the set of residue classes of integers modulo

Math. Mag. **92** (2019) 288–295. doi:10.1080/0025570X.2019.1614820
MSC: Primary 05C20, Secondary 11T06; 05C60
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

p, where each class is represented by the unique integer i, $0 \le i \le p-1$ belonging to that class.

It is known (see, for example, Ireland and Rosen [12]) that with respect to modular arithmetic, $\mathbb{Z}_p$ is a field. For instance, in $\mathbb{Z}_7$, $1+6=0$, $3\cdot 4=5$, and $3^{-1}=5$ since $3\cdot 5=1$. One can consider polynomials with coefficients from $\mathbb{Z}_p$; let $\mathbb{Z}_p[X]$ denote the set of all of them. Every polynomial $f \in \mathbb{Z}_p[X]$ defines a function on $\mathbb{Z}_p$ when it is evaluated at elements of $\mathbb{Z}_p$. For example, for $f = X^3 - 4X + 6 = X^3 + 3X + 6 \in \mathbb{Z}_7[X]$, $f(0)=6$, $f(1)=1^3+3\cdot 1+6=10=3$, and $f(2)=20=6$. Also, $f(3)=42=0$, and we say that 3 is a root of f. Counting or estimating the number of roots of polynomials from $\mathbb{Z}_p[X]$ in $\mathbb{Z}_p$ is a fundamental problem in the area of mathematics called algebraic geometry.

All results in this article hold over any finite field of odd characteristic, but for ease of presentation we will simply use the field $\mathbb{Z}_p$, $p>2$.

Here is the problem. Recently we were surprised to learn that for any prime $p \ge 3$ and any natural numbers m and n satisfying $mn \equiv 1 \bmod (p-1)$, the trinomials $X^{m+1} - 2X + 1$ and $X^{n+1} - 2X + 1$ have the same number of distinct roots in $\mathbb{F}_p$. Of course, the coefficients 1 and -2 are elements of $\mathbb{F}_p$ and $-2 = p-2$. For example, it is easy to check that for $p=11$, $m=3$, $n=7$, the trinomial $X^{m+1} - 2X + 1$ has roots 1, 5, and 8 (with root 8 of multiplicity 2), and the trinomial $X^{n+1} - 2X + 1$ has roots 1, 2, and 3.

How did we arrive at this strange fact? We will explain it a bit later, after we discuss a special class of digraphs.

A *directed graph*, or just *digraph*, $D=(V,A)$ is a pair of two sets V and $A \subseteq V \times V$; V is called the set of *vertices* of D, and A is called the set of *arcs* of D. All undefined terms related to digraphs can be found in Bang-Jensen and Gutin [4].

The digraph D of Figure 1(a) has vertex set $V=\{a,b,c,d\}$ and arc set $A=\{(a,b),(a,c),(b,b),(b,c),(c,a),(d,a),(d,c)\}$. Arc (b,b) is called a *loop*, and we say that vertex b has a loop on it.

Given a digraph $D=(V,A)$, a digraph $H=(V_1,A_1)$ is called a *subdigraph* of D if $V_1 \subseteq V$ and $A_1 \subseteq A$ (see Figure 1(b) for an example of a subdigraph).

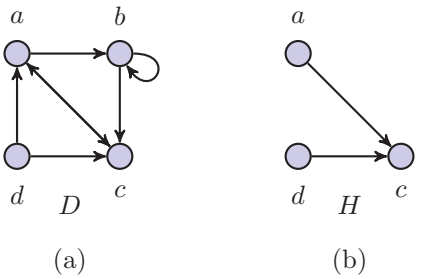


Figure 1 Digraph D and its subdigraph H.

We shall be interested in a certain type of digraphs. For any positive prime p, and any positive integers m, n, we define the directed graph $D(p;m,n)=(V,A)$, with vertex set $V=\mathbb{F}_p \times \mathbb{F}_p$ and arc set A as follows: the ordered pair of vertices $((x_1,x_2),(y_1,y_2))$ is an arc if

$$x_2 + y_2 = x_1^m y_1^n.$$

We call $D(p;m,n)$ a *monomial digraph*. It is known (and often referred to as Fermat's little theorem) that $x^p = x$ for any $x \in \mathbb{F}_p$. It is therefore sufficient to restrict

integers m and n in the definition of $D(p; m, n)$ to the set $\{1, \ldots, p-1\}$. We thus have $(p-1)^2$ digraphs $D(p; m, n)$ for every prime p.

The digraphs $D(p; m, n)$ are directed analogues (see Kodess [**15**], Kodess and Lazebnik [**16**]) of particular cases of a well studied class of algebraically defined undirected graphs having many applications; see surveys by Lazebnik and Woldar [**17**] and Lazebnik, Sun, and Wang [**18**].

Figure 2 shows $D(3; 1, 2)$. Note that $((2, 2), (1, 0))$ is an arc, since according to the adjacency condition, $2 + 0 = 2^1 \cdot 1^2$ in $\mathbb{F}_3$; $((1, 0), (2, 2))$ is not an arc, since $0 + 2 \neq 1^1 \cdot 2^2$ in $\mathbb{F}_3$; and vertex $(1, 2)$ has a loop on it since, $2 + 2 = 1 \cdot 1^2$ in $\mathbb{F}_3$.

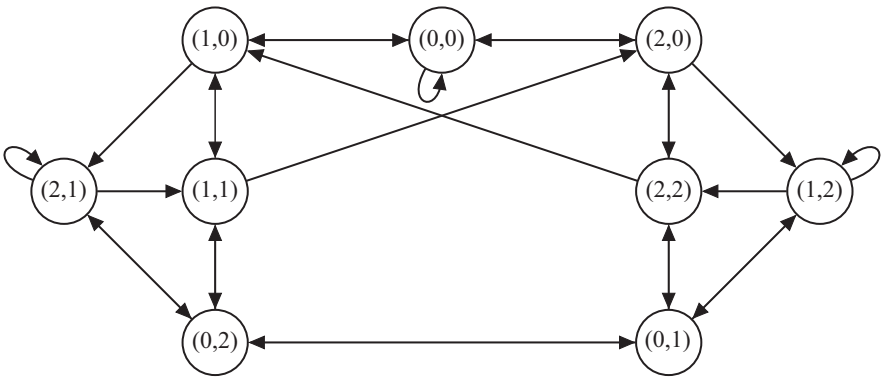


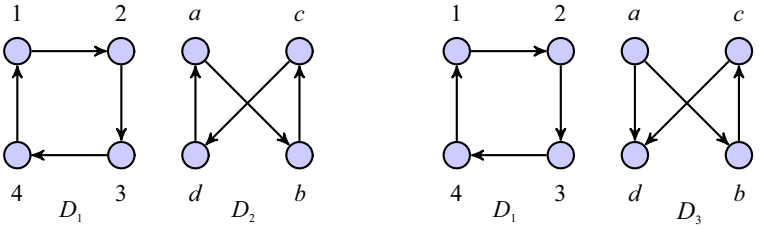
Figure 2 The digraph $D(3; 1, 2)$: $x_2 + y_2 = x_1 y_1^2$.

As all these $(p-1)^2$ digraphs $D(p; m, n)$ share the same vertex set, one cannot help wondering if they are actually different. For instance, it is not hard to see that $D(3; 1, 2)$ and $D(3; 2, 1)$ can be obtained from one another by reversing the orientation of every arc. The reason for this will become more clear later. A very thorough and tedious inspection (perhaps done by computer) would reveal that the digraphs $D(5; 1, 2)$ and $D(5; 3, 2)$, both having 25 vertices, are in fact not much different: one can be obtained from the other by relabeling the vertices.

We would like to make this discussion a little more formal.

Central to many areas of mathematics is the concept of *isomorphism*. It is defined for such ubiquitous and important objects as vector spaces, groups, fields and graphs, to name just a few. Informally, two objects are called isomorphic if they are not fundamentally different in their structure; that is, one of them can be obtained from the other by renaming or relabeling the elements while preserving the internal structure. Formally, we call digraphs D_1 and D_2 isomorphic and write $D_1 \cong D_2$ if there is a bijective function f from the vertex set $V(D_1)$ of D_1 to the vertex set $V(D_2)$ of D_2 such that for any two vertices $u, v \in V(D_1)$, (u, v) is an arc in D_1 if and only if $(f(u), f(v))$ is an arc in D_2. That is, f preserves adjacency and non-adjacency. Such a mapping f is called an *isomorphism from* D_1 *to* D_2.

To illustrate the idea of a graph isomorphism, we refer to Figure 3. The two digraphs on the left of the figure, D_1 and D_2, are isomorphic because the mapping defined as $f(1) = a$, $f(2) = b$, $f(3) = c$, $f(4) = d$, is clearly a bijection; and it is a routine verification to check that f preserves adjacency and non-adjacency. For instance, $(2, 3)$ is an arc in D_1, and its image $(f(2), f(3)) = (b, c)$ is an arc in D_2, whereas $(1, 3)$ and its image $(f(1), f(3)) = (a, c)$ are not arcs in D_1 and D_2, respectively.

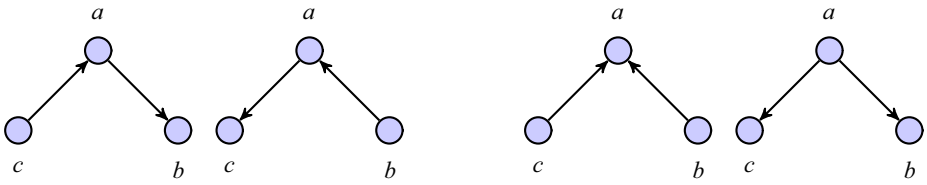


(a) Two isomorphic digraphs.

(b) Two non-isomorphic digraphs.

Figure 3 The concept of isomorphism of digraphs.

The reader should be convinced that not only arcs but all digraph-theoretic properties (that is, those independent of the actual labeling of the vertices) are shared by two isomorphic digraphs. For example, if g is an isomorphism from a digraph H_1 to a digraph H_2, then every vertex x of H_1 and the vertex $g(x)$ of H_2 have the same number of in-going arcs and the same number of out-going arcs. This observation helps establishing the fact that the two digraphs on the right in Figure 3, D_1 and D_3, are not isomorphic: in D_3 vertex a has two out-going arcs, whereas D_1 has no vertex with this property. Other properties shared by isomorphic (di)graphs include the number of (directed) cycles of a given length, the total number of (directed) cycles, the number of (strong) components, etc. Note that simply reversing the orientation of every arc in a digraph may or may not produce a digraph isomorphic to the original one. See Figure 4.



(a) Reversing the arcs produces an isomorphic digraph.

(b) Reversing the arcs produces a non-isomorphic digraph.

Figure 4 Reversing the arcs of a digraph.

Suppose one has a large set of digraphs and wants to find all its members with a particular property. Every member of the set can be considered and checked for having the property, but, as isomorphic digraphs possess the property simultaneously, it is sufficient to check only one of them. So only one member of a class of isomorphic digraphs can be considered. Therefore, if one has an efficient way for establishing isomorphism of digraphs from the set, the original set of digraphs can be replaced by a smaller subset of it (and often much smaller) consisting of one "representative" of each class of isomorphic graphs, and the property can be checked only for digraphs from this subset. This approach becomes even more efficient when we wish to check multiple properties for digraphs from the original set. Once its members are "sorted for isomorphism", every property can be checked for only one representative of each class. Unfortunately, establishing an isomorphism between digraphs is often not easy.

Asking whether two objects are isomorphic and searching for effective computational tools for answering this question has been the subject of a number of highly

publicized mathematical endeavors in the 20th century. The reader may have heard of the classification of finite simple groups problem that seeks to give a complete list of such groups up to isomorphism; see expositions by Solomon [21,22]. Another example is the graph isomorphism problem which is concerned with finding fast algorithms for determining whether two finite graphs are isomorphic.

The question of whether there is an isomorphism between two monomial digraphs $D_1 = D(p; m_1, n_1)$ and $D_2 = D(p; m_2, n_2)$ is open, and it is this question that originally motivated us. In an attempt to answer this question one would seek necessary and sufficient conditions on the parameters m_1, n_1, m_2, n_2 under which the two digraphs D_1 and D_2 are isomorphic. One idea to attack this problem is to look at various subdigraphs of D_1 and D_2.

Let X and Y be arbitrary digraphs, and let $N(X, Y)$ denote the number of subdigraphs of X each of which is isomorphic to Y. In trying to decide whether two given digraphs X_1 and X_2 are isomorphic one could hope to find a "test digraph" Y such that $N(X_1, Y) = N(X_2, Y)$ if and only if $X_1 \cong X_2$. This approach was successful in the case of a certain class of undirected graphs, see Dmytrenko, Lazebnik, and Viglione [8]. In attempting to replicate this success for the class of monomial digraphs, we were led to consider the digraph K of Figure 5. We must admit at this point that K was not a good test digraph: much to our regret, we discovered a great many pairs of non-isomorphic monomial digraphs D_1 and D_2 containing the same number of (isomorphic) copies of K. Counting $N(D(p; m, n), K)$, however, led to the result on the number of roots of certain polynomials over finite fields that we have already mentioned. Let us present our solution.

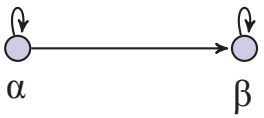


Figure 5 The digraph K.

Theorem. *For any odd prime p and any natural numbers m and n satisfying $mn \equiv 1 \bmod (p-1)$, the trinomials $X^{m+1} - 2X + 1$ and $X^{n+1} - 2X + 1$ have the same number of distinct roots in $\mathbb{F}_p$.*

Proof. Set $D_m = D(p; 1, m)$, $D_n = D(p; 1, n)$ and $D'_n = D(p; n, 1)$. We first argue that D_m and D'_n are isomorphic, and as such, contain the same number of isomorphic copies of K shown in Figure 5.

As $mn \equiv 1 \bmod (p-1)$ implies $n \cdot m - t \cdot (p-1) = 1$ for some integer t, we conclude that $\gcd(m, p-1) = 1$.

We now recall that the multiplicative group $\mathbb{F}_p^*$ of $\mathbb{F}_p$ is cyclic of order $p-1$. This implies by the elementary theory of cyclic groups that $x \mapsto x^m$ is a permutation (bijective function) on $\mathbb{F}_p$. Also for any integers k, l and any $x \in \mathbb{F}_p$, $k \equiv l \bmod (p-1)$ implies $x^k = x^l$. Proofs of these facts can be found in [12].

Consider the mapping $\psi : V(D_m) \to V(D'_n)$ defined by $\psi((x, y)) = (x^m, y)$. We verify that ψ satisfies the definition of digraph isomorphism. Clearly ψ is a permutation on $\mathbb{F}fp^2 = V(D_m) = V(D'_n)$. Suppose $((x_1, x_2), (y_1, y_2))$ is an arc in D_m, that is,

$$x_2 + y_2 = x_1 y_1^m.$$

Then its image $(\psi((x_1, x_2)), \psi((y_1, y_2))) = ((x_1^m, x_2), (y_1^m, y_2))$ is an arc in D'_n since

$$x_2 + y_2 = x_1^1 y_1^m = x_1^{mn} y_1^m = (x_1^m)^n (y_1^m)^1.$$

Similarly, we show that ψ preserves non-adjacency: if $((x_1, x_2), (y_1, y_2))$ is not an arc in D_m, then

$$x_2 + y_2 \neq x_1^1 y_1^m,$$

and so $(\psi((x_1, x_2)), \psi((y_1, y_2)))$ is not an arc in D'_n. This implies by definition that D_m and D'_n are isomorphic. Hence, $N(D_m, Y) = N(D'_n, Y)$ for any digraph Y. In particular, $N(D_m, K) = N(D'_n, K)$.

Now let H^c denote the *converse* of digraph H, that is, the digraph obtained from H by reversing all its arcs. Obviously, for any digraph D, $N(D, H) = N(D^c, H^c)$, and also $(H^c)^c = H$. Observe that D'_n is simply D_n with all arcs reversed, that is $D'_n = D_n^c$. Thus $D_n'^c$ and $(D_n^c)^c = D_n$ are equal. Since $K^c \cong K$, we have

$$N(D_m, K) = N(D'_n, K) = N(D_n'^c, K^c) = N(D_n'^c, K) = N(D_n, K).$$

Note that we did not assume that D_m and D_n were isomorphic! Actually we conjecture that they never are unless $m = n$.

We now show that the number of isomorphic copies of K in a digraph D_n can be expressed as the number of distinct roots of a polynomial of degree $n + 1$ in the field $\mathbb{F}_p$.

Suppose K is a subgraph of $D_n = D(p; 1, n)$, and let $\alpha = (u, s)$ and $\beta = (v, t)$ be vertices of K. From the relations defining the three arcs of K, we have

$$s + s = u \cdot u^n, \quad t + t = v \cdot v^n, \quad \text{and} \quad s + t = uv^n.$$

Note that since p is odd, $2 \in \mathbb{F}_p$ is invertible. Hence, we obtain

$$s = \frac{1}{2}u^{n+1}, \quad t = \frac{1}{2}v^{n+1}, \quad \text{and } s + t = uv^n. \tag{1}$$

If $u = v$, then the first and the second equation of system (1) imply $s = t$, and so vertices α and β are equal. Therefore, $u \neq v$.

It follows from (1) that the equation $s + t = uv^n$ can be rewritten as

$$\frac{1}{2}u^{n+1} + \frac{1}{2}v^{n+1} = uv^n. \tag{2}$$

Note that neither u nor v is 0. Indeed, if $u = 0$, then substituting it to the first and to the third equation of the system 1, we get $s = 0$ and $s + t = 0$. Hence, $t = 0$, and from the second equation we get $v = 0$. Hence, $\alpha = \beta = (0, 0)$—a contradiction. Therefore, $u \neq 0$. Similarly, $v \neq 0$, and $uv \neq 0$. Dividing both sides of equation 2 by $(1/2)uv^n$, we obtain

$$(u/v)^n + (v/u) = 2.$$

Setting $w = u/v$, we rewrite this equation as $w^{n+1} - 2w + 1 = 0$. Hence, u/v is a root of the polynomial $f_n(X) = X^{n+1} - 2X + 1 \in \mathbb{F}_p[X]$. It is not equal to the obvious root 1, as $u \neq v$.

Consequently, $N(D_n, K) = (p - 1)R(n)$, where $R(n)$ is the number of distinct roots of f_n in $\mathbb{F}_p \setminus \{1\}$; any choice of root and any choice of u must determine α and β uniquely.

Now if m and n are integers satisfying the conditions of the theorem, we have from before that $(p - 1)R(m) = N(D_m, K) = N(D_n, K) = (p - 1)R(n)$, and so $R(m) = R(n)$. ■

Thus, from an isomorphism problem for digraphs, we have arrived at an interesting fact concerning trinomials over finite fields. The theorem can be immediately generalized in various ways and proved directly, i.e., without considering graphs or digraphs. We suggest that the reader find a proof for the following generalization.

Exercise. For any prime power q (even or odd) and any natural numbers m and n satisfying $mn \equiv 1 \bmod (q-1)$, polynomials $X^{m+1} + aX + b$ and $X^{n+1} + aX + b^m$ have the same number of distinct roots in the finite field $\mathbb{F}_q$ for any $a, b \in \mathbb{F}_q$.

We end this note with two open questions concerning monomial digraphs which we find very interesting. Though we do not know the answers even for prime q, we state the questions for any prime power q. Let $D_1 = D(q; m_1, n_1)$ and $D_2 = D(q; m_2, n_2)$.

Problem 1. Is there a digraph H such that the equality $N(D_1, H) = N(D_2, H)$ is equivalent to $D_1 \cong D_2$?

Problem 2. Find necessary and sufficient conditions on q, m_1, n_1, m_2, n_2, such that digraphs D_1 and D_2 are isomorphic.

A related conjecture appears in [15]:

Conjecture. *Let q be a prime power, and let m_1, n_1, m_2, n_2 be integers from $\{1, 2, \ldots, , q-1\}$. Then $D(q; m_1, n_1) \cong D(q; m_2, n_2)$ if and only if there exists an integer k, coprime with $q-1$ such that*

$$m_2 \equiv km_1 \mod (q-1) \text{ and } n_2 \equiv kn_1 \mod (q-1).$$

Acknowledgments The authors are thankful to the anonymous referees whose thoughtful comments improved the paper. The work of the last author was partially supported by a grant from the Simons Foundation #426092.

REFERENCES

[1] Alon, N. (2002). Discrete Mathematics: Methods and Challenges. *Proceedings of the International Congress of Mathematicians*, Vol. I (Beijing, 2002), 119–135, Beijing, China: Higher Ed. Press.
[2] Alon, N., Spencer, J. (2016). *The Probabilistic Method*. 4th ed. Hoboken, NJ: Wiley.
[3] Babai, L., Frankl, P. (1992). *Linear Algebra Methods in Combinatorics*. Department of Computer Science, University of Chicago, preliminary version.
[4] Bang-Jensen, J. Gutin, G. (2009). *Digraphs. Theory, Algorithms and Applications*. 2nd ed., Springer Monographs in Mathematics, New York, NY: Springer-Verlag.
[5] Biggs, N. (1994). *Algebraic Graph Theory*. 2nd ed., Cambridge, UK: Cambridge University Press.
[6] Bondarenko, A. (2014). On Borsuk's conjecture for two-distance sets. *Discrete & Computational Geometry* 51(3): 509–515.
[7] Brouwer, A. E., Haemers, W. H. (2012). *Spectra of Graphs*. New York: Springer-Verlag.
[8] Dmytrenko, V., Lazebnik, F., Viglione, R. (2005). Isomorphism criterion for monomial graphs. *J. Graph Theory* 48: 322–328.
[9] Doob, M. (1984) Applications of graph theory in linear algebra. *Math. Mag.*, 57(2): 67–76.
[10] Erdős, P., Meir, A., Sos V.T., Turán, P. (1972). On some applications of graph theory, III. *Canad. Math. Bull.* 15: 27- 32.
[11] Füredi, Z. Simonovits, M. (2013). The history of degenerate (bipartite extremal graph problems. *Erdős centennial*, 169–264, Bolyai Soc. Math. Stud., vol. 25, Budapest: Janos Bolyai Math. Soc.
[12] Ireland, K., Rosen, M. (1990). *A Classical Introduction to Modern Number Theory*. New York, NY: Springer.
[13] Kahn, J., Kalai, G. (1993). A counterexample to Borsuk's conjecture. *Bull. Am. Math. Soc.* 29(1): 60–62.
[14] Katona, G. (1969). Graphs, vectors and inequalities in probability theory. *Mat. Lapok* 20: 123–127.
[15] Kodess, A. (2014). Properties of some algebraically defined digraphs. Ph.D. thesis. University of Delaware, Newark, DE.

[16] Kodess, A., Lazebnik, F. (2015). Connectivity of some algebraically defined digraphs. *Electron. J. Combin.* 22(3): Paper 27, 11pp.
[17] Lazebnik, F., Woldar, A. J. (2001). General properties of some families of graphs defined systems of equations. *J. Graph Theory*, 38: 65–86.
[18] Lazebnik, F., Sun, S., Wang, Y. (2017). Some families of graphs, hypergraphs and digraphs defined by systems of equations: A survey. *Lecture Notes of Seminario Interdisciplinare di Matematica* 14: 105–142.
[19] Lovász, L. (1978). Kneser's conjecture, chromatic number, and homotopy. *J. Combin. Th.*, 25: 319–324.
[20] Lovász, L., Pyber, L., Welsh, D. J. A., Ziegler, G. M. (1995). Combinatorics in pure mathematics. In: *Handbook of combinatorics*, Graham, R.L., Grötschel, M., Lovász, L. eds. Cambridge, MA: MIT Press, 2039–2082.
[21] Solomon, R. (1995). On finite simple groups and their classification. *Notices Amer. Math. Soc.* 42(2): 231–239.
[22] Solomon, R. (2001). A brief history of the classification of the finite simple groups. *Bull. Amer. Math. Soc.* (N.S.) 38(3): 315–352.
[23] Swan, R. G. (1963/1969). An application of graph theory to algebra. *Proc. Amer. Math. Soc.* 14: 367–373; Correction to "An application of graph theory to algebra". *Proc. Amer. Math. Soc.* 21: 379–380.
[24] Turán, P. (1970). Applications of graph theory to geometry and potential theory. In: Guy, R.K., Hanani, H., Sauer, N., and Schonheim, J., *Combinatorial Structures and their Application*. New York: Gorgon and Breach, pp. 423–434.

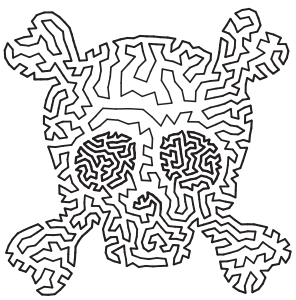
Summary. We present an example of a result in graph theory that is used to obtain a result in another branch of mathematics. More precisely, we show that the isomorphism of certain directed graphs implies that some trinomials over finite fields have the same number of roots.

ROBERT S. COULTER (MR Author ID: 615822) joined the faculty at the University of Delaware in 2003, having previously held positions at the University of Queensland, Queensland University of Technology, and Deakin University. He received his Ph.D. from the Department of Computer Science and Electrical Engineering at the University of Queensland in 1998. He is an Australian, and greatly misses Farmers Union Iced Coffee and the lack of snow shovels.

STEFAAN DE WINTER (MR Author ID: 723701) moved to Michigan Technological University in 2011 and made such a great impression that he was promoted early to Associate Professor. Though he heralds from Belgium, he abandoned the great chocolate and beer of his homeland, not to mention the cycling, and moved to the United States in pursuit of happiness, a pursuit in which he was ironically successful through the medium of another foreign national!

ALEX KODESS (MR Author ID: 886420) has recently moved to the Empire State and joined the Mathematics Department of Farmingdale State College. In the past he escaped the clutches of the state institution of the second smallest state in the country, only to be subsumed by the smallest. In exchanging the University of Rhode Island for the University of Delaware, he could at least be content in the knowledge his status "improved" from Ph.D. student to faculty member. His only regret is that he now finds he is expected to act like a responsible adult.

FELIX LAZEBNIK (MR Author ID: 111260) has been at the University of Delaware since receiving his Ph.D. from the University of Pennsylvania under Herbert S. Wilf in 1987. He claims to understand Robert's English, Alex's Ph.D. Thesis and some of Stefaan's geometry.



Opt-emoji Skull and Crossbones, Robert Bosch; digital print, 2017. An optimal TSP tour of 1024 points that were arranged to resemble an emoji rendition of a skull and crossbones symbol. The optimal tour was obtained with the Concorde TSP Solver.
See interview on page 305.

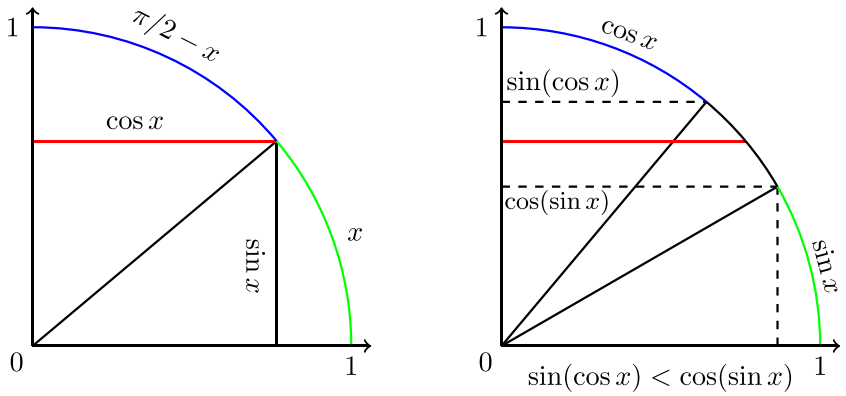
Proof Without Words: $\sin(\cos x) < \cos(\sin x)$

VINCENT FERLINI
Keene State College
Keene, NH 03435
vferlini@keene.edu

To show that $\sin(\cos x) < \cos(\sin x)$ for all x, it suffices to establish the inequality for $x \in (0, \pi/2)$ for the following reasons:

1) by periodicity we need only consider $x \in [-\pi, \pi]$;
2) $\sin(\cos x)$ and $\cos(\sin x)$ are even functions so we need only consider $x \in [0, \pi]$;
3) $\sin(\cos x) < \cos(\sin x)$ for $x \in (-\pi/2, \pi]$ and so we need only consider $x \in [0, \pi/2]$; and
4) $\sin(\cos 0) = \sin 1 < 1 = \cos(\sin 0)$ and $\sin(\cos(\pi/2)) = 0 < \cos 1 = \cos(\sin(\pi/2))$ so we need only consider $x \in (0, \pi/2)$.

A wordless proof of $\sin(\cos x) < \cos(\sin x)$ for $x \in (0, \pi/2)$ appears below.



Summary. The trigonometric inequality $\sin(\cos x) < \cos(\sin x)$ is proved visually for $x \in (0, \pi/2)$.

VINCENT FERLINI (MR Author ID: 673403) is a Professor of Mathematics at Keene State College. His interests lie mainly in group theory, number theory, and geometry.

Math. Mag. **92** (2019) 296. doi:10.1080/0025570X.2019.1637678
MSC: Primary 51N20
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

The Delian Problem, Platonic Solids, and Finite Fields

MATT D. LUNSFORD
Union University
Jackson, TN 38305
mlunsford@uu.edu

The question of performing, by straightedge and compass only, the classical Greek construction of duplicating a cube remained a significant open problem in mathematics until the first half of the 19th century. Recall that the problem of duplicating a cube is to construct a cube whose volume is exactly twice the volume of a given cube. We assume, following Plato, that the construction must be performed using only a straightedge and collapsible compass, and that the given object, the cube to be duplicated, is indeed constructible. Traditionally, this problem has been associated with the island of Delos and allegedly results from the desire to double the volume of a cube-shaped altar to the god Apollo. Accordingly, this geometric construction has been labeled as the Delian problem.

In 1837, Pierre Wantzel [1] proved that the duplication of a cube is impossible. Wantzel's argument translates the geometric construction into an algebraic characterization and employs the algebraic properties of constructible numbers and the irreducibility of a specific cubic polynomial. Before recalling the details, we begin with a review of the necessary essentials.

Let F be a field. E is an extension field of F if E is also a field and $F \subseteq E$. The vector space dimension of E over F is denoted by $[E : F]$ and is called the degree of the field extension. If K is an intermediate field such that $F \subseteq K \subseteq E$, then $F \subseteq K \subseteq E$ is called a tower of fields. Moreover, $[E : F] = [E : K][K : F]$. Hence, $[K : F]$ is a divisor of $[E : F]$.

A polynomial $f(x) \in F[x]$ is *irreducible* over F if it cannot be written as the product of two polynomials in $F[x]$ of lower degree. An element a is algebraic over F if a is the root of some polynomial $f(x) \in F[x]$. A real number a, which is algebraic over the field $\mathbb{Q}$ of rational numbers, is said to be a constructible number if there exist a tower of fields

$$\mathbb{Q} = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t \subseteq \mathbb{R},$$

such that $a \in K_t$ and $[K_{i+1} : K_i] = 1$ or 2 for $i = 0, \ldots, t-1$, where $\mathbb{R}$ is the field of real numbers.

Wantzel observed that the ability to duplicate a cube is equivalent to the constructibility of the real number $\sqrt[3]{2}$. Using the fact that the cubic polynomial $x^3 - 2$ is irreducible over $\mathbb{Q}$, he deduced that $\left[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}\right] = 3$. It follows then from the definition above that $\sqrt[3]{2}$ is not a constructible number. Therefore, the geometric construction of duplicating a cube by straightedge and compass alone is impossible.

In this paper, we first consider the question of duplication for the other four Platonic solids. Then, we exchange the geometric notions of constructible number and duplication of a Platonic solid with their algebraic counterparts and seek results over finite fields.

Math. Mag. **92** (2019) 297–299. doi:10.1080/0025570X.2019.1637677
MSC: Primary 12F05

The volume formulas for the Platonic solids, the cube, tetrahedron, octahedron, dodecahedron, and icosahedron, can be expressed in terms of the edge length a of the solid [2, p.127], respectively, as

$$V_{\text{cube}} = c_0 a^3,\ V_{\text{tet}} = c_1 a^3,\ V_{\text{oct}} = c_2 a^3,\ V_{\text{dod}} = c_3 a^3,\ \text{and}\ V_{\text{ico}} = c_4 a^3,$$

where

$$c_0 = 1,\ c_1 = \frac{\sqrt{2}}{12},\ c_2 = \frac{\sqrt{2}}{3},\ c_3 = \frac{15+7\sqrt{5}}{4},\ \text{and}\ c_4 = \frac{15+5\sqrt{5}}{12}.$$

Using Wantzel's idea, one easily proves that the duplication of any Platonic solid by straightedge and compass alone is impossible. In fact, for $1 \le i \le 4$, the minimal polynomial for $\sqrt[3]{2c_i}$ is of degree six and thus $\left[\mathbb{Q}(\sqrt[3]{2c_i} : \mathbb{Q}\right] = 6$. Hence, the edge length required for duplication is not a constructible number.

Let's move on to the finite field case. For the remainder of this paper, p will denote an odd prime, F_p will denote the prime finite field of characteristic p, and $F_p^* = F_p \setminus \{0\}$. Now, we extend the notion of constructible number to prime finite fields.

Definition 1. An element a, which is algebraic over F_p, is said to be a *purely quadratic* element over F_p if there exists a tower of fields $F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t$ such that $a \in K_t$ and $[K_{i+1} : K_i] = 1$ or 2 for $i = 0, \ldots, t-1$.

We are particularly interested in prime finite fields for which a cube root of $2c_i$ is a purely quadratic element. Let's begin with a cube root of 2.

Definition 2. F_p is a Delian field if a cube of root of 2 is a purely quadratic element over F_p.

The following lemma moves the discussion from purely quadratic elements to the existence of cube roots in the prime finite field.

Lemma. *F_p is Delian if and only if the cubic polynomial $x^3 - 2$ has a root in F_p.*

Proof. Let α denote a cube root of 2. Then $[F_p(\alpha) : F_p] \le 3$. From the definition above, α will be purely quadratic over F_p exactly when the degree of this extension does not equal 3. This, in turn, is equivalent to the reducibility of $x^3 - 2$ and therefore the existence of a root of $x^3 - 2$ in F_p. ■

We can now state the key result of the paper.

Theorem. *F_p is Delian if and only if one of the following holds:*

i) $p \equiv 0$ *or* $2 \bmod 3$, *or*

ii) $p \equiv 1 \bmod 3$ *and* $p = a^2 + 27b^2$ *with* $a, b \in \mathbb{N}$.

Proof. If $p = 3$, then Fermat's little theorem guarantees a cube root of 2 in F_3. If $p \equiv 2 \bmod 3$, then the map $c \to c^3$ on F_p^* is injective since F_p^* has no element of order 3. Hence, the mapping is in fact a bijection, and it follows that 2 is a cube root in F_p. Finally, if $p \equiv 1 \bmod 3$, then the cubic reciprocity theorem of Gauss [3, p.72] states that 2 is a cube root in F_p if and only if $p = a^2 + 27b^2$. ■

This theorem completely characterizes Delian fields and answers the question of when it is possible to "algebraically" duplicate a cube over a prime finite field.

We now move to the duplication of other Platonic solids over prime finite fields. Consider the field extension $GF(p^2)$ of degree 2 over F_p, with $p > 5$. As all quadratics in $F_p[x]$ are reducible over $GF(p^2)$, we know that both a square root of 2 and a square

root of 5 lie in this extension field. Moreover, we know for sure that each c_i and, in particular, each $2c_i$ exists and is nonzero. Furthermore, it is evident that a cube root of $2c_i$ is a purely quadratic element over F_p if and only if it lies in $GF(p^2)$. The following examples demonstrate that this necessary and sufficient condition must be investigated for each odd prime $p > 5$ and for each $i = 1, 2, 3, 4$. First, we make the following definition.

Definition 3. F_p is a *Platonic* field if a cube root of $2c_i$ is a purely quadratic element over F_p, for $0 \leq i \leq 4$.

Examples. F_{11} is a Platonic field. In particular, $\sqrt[3]{2c_i}$ is purely quadratic over F_{11}, $0 \leq i \leq 4$, since all cube roots of $2c_i$ already reside in $GF(11^2)$. To see this, let β be a square root of 2 in $GF(11^2)$. Then,

$$\sqrt[3]{2c_0} = 7,\ \sqrt[3]{2c_1} = \beta,\ \sqrt[3]{2c_2} = 5\beta,\ \sqrt[3]{2c_3} = 3,\ \text{and}\ \sqrt[3]{2c_4} = 5$$

are all in $GF(11^2)$.

In contrast, $\sqrt[3]{2c_i}$, $0 \leq i \leq 4$, is not purely quadratic over F_7, since all cube roots of $2c_i$ reside in the field extension $GF(7^6)$ of degree 3 over $GF(7^2)$. Let γ be a square root of 5 in $GF(7^2)$. Then, the elements

$$\sqrt[3]{2c_0} = \sqrt[3]{2},\ \sqrt[3]{2c_1} = \sqrt[3]{4},\ \sqrt[3]{2c_2} = \sqrt[3]{2},\ \sqrt[3]{2c_3} = \sqrt[3]{4},\ \text{and}\ \sqrt[3]{2c_4} = \sqrt[3]{6 + 2\gamma}$$

are not present in $GF(7^2)$.

Finally, to demonstrate non-extreme possibilities, we note that if $p = 13$, then $\sqrt[3]{2c_i}$ is purely quadratic over F_{13} when $i = 1, 3$, but is not purely quadratic when $i = 0, 2, 4$.

Current citizens of Delos should be pleased by the fact that even though the duplication of a cube (or any Platonic solid) cannot be accomplished geometrically, it can be accomplished "in a sense" algebraically using purely quadratic elements over a Delian (or Platonic) field.

REFERENCES

[1] Wantzel, P. (1837). Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas. *J. Math. Pures. Appl.* 2: 366–372.
[2] Beyer, W. H. (1984). *CRC Standard Mathematical Tables*. Boca Raton, FL: CRC Press.
[3] Cox, D. A. (2013). *Primes of the Form $x^2 + ny^2$*. Hoboken, NJ: John Wiley & Sons, Inc.

Summary. The question of performing, by straightedge and compass only, the classic Greek construction problem of duplicating a cube remained a significant open problem in mathematics until the first half of the 19th century. By using Wantzel's algebraic characterization of the Delian problem, we extend the notion of being constructible to Platonic solids over prime finite fields.

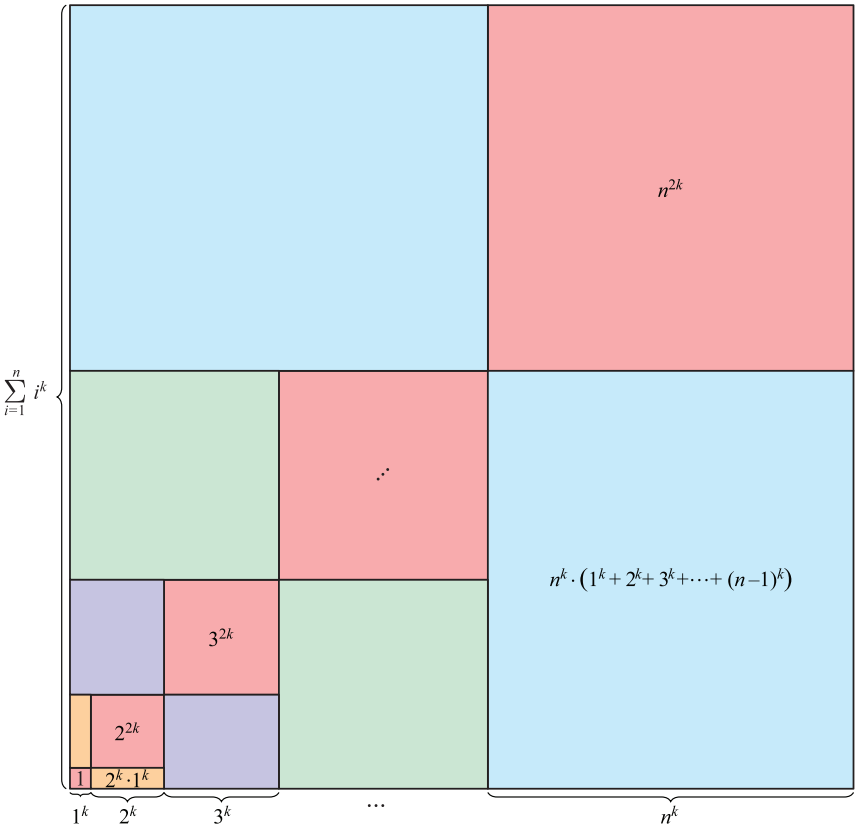
MATT D. LUNSFORD (MR Author ID: 367743) is Professor of Mathematics at Union University in Jackson, TN, where he has been a faculty member since 1993. He holds a doctorate in mathematics from Tulane University. His current research interests include classical Galois theory and finite fields. He and his wife Deanna have three children: Cara, Thomas, and Emma, and one son-in-law: Brennan.

Proof Without Words: Sums of Even and Odd Powers

TOM EDGAR
Pacific Lutheran University
Tacoma, WA 98447
edgartj@plu.edu

Chilaka [1], Markham [2], and Yates [3] provide nearly identical diagrams to visually compute the partial sums $\sum_{i=1}^{n} i^2$ and $\sum_{i=1}^{n} i^4$. We generalize their diagrams to demonstrate a visual computation of any partial sum of the form $\sum_{i=1}^{n} i^{2k}$ where $k \geq 1$ is an integer.

Theorem. *For any integer $k \geq 1$, we have* $$\sum_{i=1}^{n} i^{2k} = \left(\sum_{i=1}^{n} i^k\right)^2 - 2\left(\sum_{j=2}^{n} j^k \sum_{i=1}^{j-1} i^k\right).$$



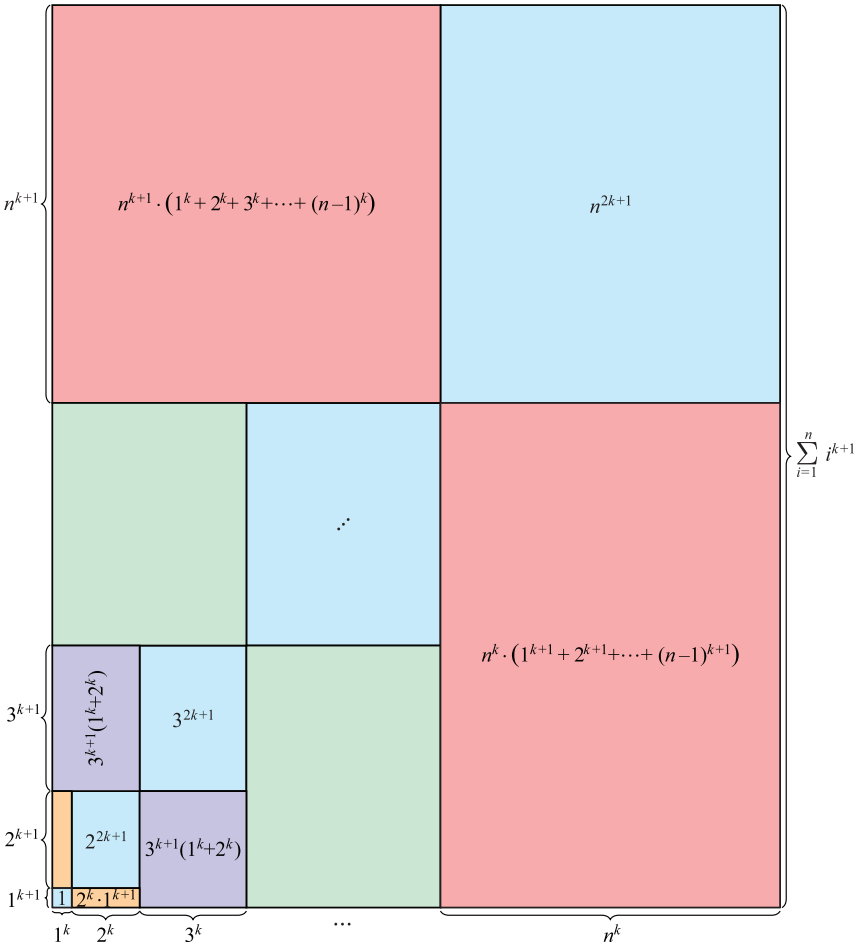
Remark. Substitution of known formulas for sums of powers (less than $2k$) yields a closed formula for $\sum_{i=1}^{n} i^{2k+1}$.

Math. Mag. **92** (2019) 300–301. doi:10.1080/0025570X.2019.1638215
MSC: Primary 05A19, Secondary 11P99
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

An anonymous referee suggested attempting to find an analogous formula and visual proof for partial sums of the form $\sum_{i=1}^{n} i^{2k+1}$ where $k \geq 0$ is an integer, which we do below.

Theorem. *For any integer $k \geq 0$, we have*

$$\begin{aligned}\sum_{i=1}^{n} i^{2k+1} &= \left(\sum_{i=1}^{n} i^{k}\right) \cdot \left(\sum_{i=1}^{n} i^{k+1}\right) - \sum_{j=2}^{n} j^{k} \sum_{i=1}^{j-1} i^{k+1} - \sum_{j=2}^{n} j^{k+1} \sum_{i=1}^{j-1} i^{k} \\ &= \left(\sum_{i=1}^{n} i^{k}\right) \cdot \left(\sum_{i=1}^{n} i^{k+1}\right) - \sum_{j=2}^{n} j^{k} \left(\sum_{i=1}^{j-1} i^{k+1} + j i^{k}\right).\end{aligned}$$



REFERENCES

[1] Chilaka, J. O. (1983). Proof without words: Sum of squares. *Math. Mag.* 56(2): 90.
[2] Markham, E. M. (1992). Proof without words. *Math. Mag.* 65(1): 55.
[3] Yates, R. C. (1959). Sums of powers of integers. *Math. Teach.* 52(4): 268–271.

Summary. We generalize two well known visualizations about sums of squares and sums of fourth powers.

TOM EDGAR (MR Author ID: 821633) is an associate professor of mathematics. He realized this generalization while teaching an inquiry based seminar course about proofs without words with participants Miguel Amezola, Yajun An, Hannah Bortel, Seth Chapman, Paul Dalenberg, Egan Dunning, Matthew Fosmire, Collin Greer, Emily Hanna, Liz Holm, Robert Jogerst, Quinton Teas, Trang Than, and Devin Tracy.

Proof Without Words: Revisiting Two Trigonometric Figures and Two Identities from Bressieu and Fincke

REX H. WU
New York Presbyterian Lower Manhattan Hospital
New York, NY 10038
rexhwu@yahoo.com

Nelsen's proof [8] of the following identity

$$\tan\theta + \sec\theta = \tan\left(\frac{\pi}{4} + \frac{\theta}{2}\right) \tag{1}$$

prompts this revisit of the two figures (reproduced in Figure 1) from Thomas Fincke [5], a sixteenth century Danish physician and mathematician who first described this identity and coined the terms *tangent* and *secant*.

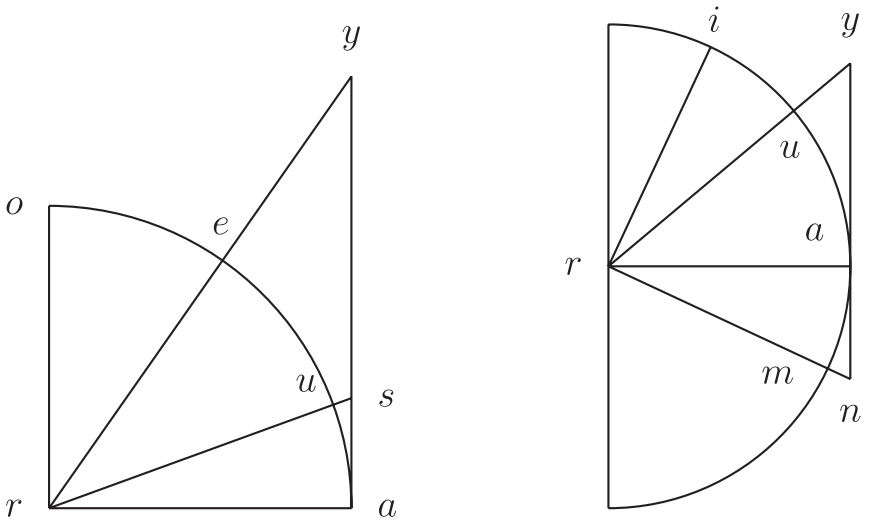


Figure 1 Arc *au* is the arc of interest. The line segments *re* and *ri* bisect the complement of arc *au* for the figure on the left and the right, respectively. For the figure on the right, arc *am* equals arc *ui*.

The posthumous publication of Regiomontanus' *De triangulis omnimodis libri quinque* in 1533 laid the foundation for the development of trigonometry in Europe. During the late sixteenth century and early seventeenth century, to assist the computation of trigonometric tables, mathematicians explored various relationships among the trigonometric functions [6]. Identities (1) through (4) are tailored for the construction of the tangent and secant tables; Identities (2) through (4) appear below:

$$\sec\theta - \tan\theta = \tan\left(\frac{\pi}{4} - \frac{\theta}{2}\right) \tag{2}$$

Math. Mag. **92** (2019) 302–304. doi:10.1080/0025570X.2019.1603732.
MSC: Primary 97G60, Secondary 01A40

$$2\sec\theta = \tan\left(\frac{\pi}{4}+\frac{\theta}{2}\right)+\tan\left(\frac{\pi}{4}-\frac{\theta}{2}\right) \tag{3}$$

$$2\tan\theta = \tan\left(\frac{\pi}{4}+\frac{\theta}{2}\right)-\tan\left(\frac{\pi}{4}-\frac{\theta}{2}\right) \tag{4}$$

In the past, trigonometric functions were not associated with angles. Rather, they were simply line segments associated with the corresponding arc. Therefore, in Figure 1 (left), the "tangent of arc au" refers to the line segment as and its "secant" is rs. Here is how Fincke stated Identity (1) (translated from Latin), "secant of an arc less than half the quadrant of a circle, with the tangent of the same (arc), equals to the tangent of the given arc and half of its complement." It is Rheticus [**9**] who freed us from the arc-function association.

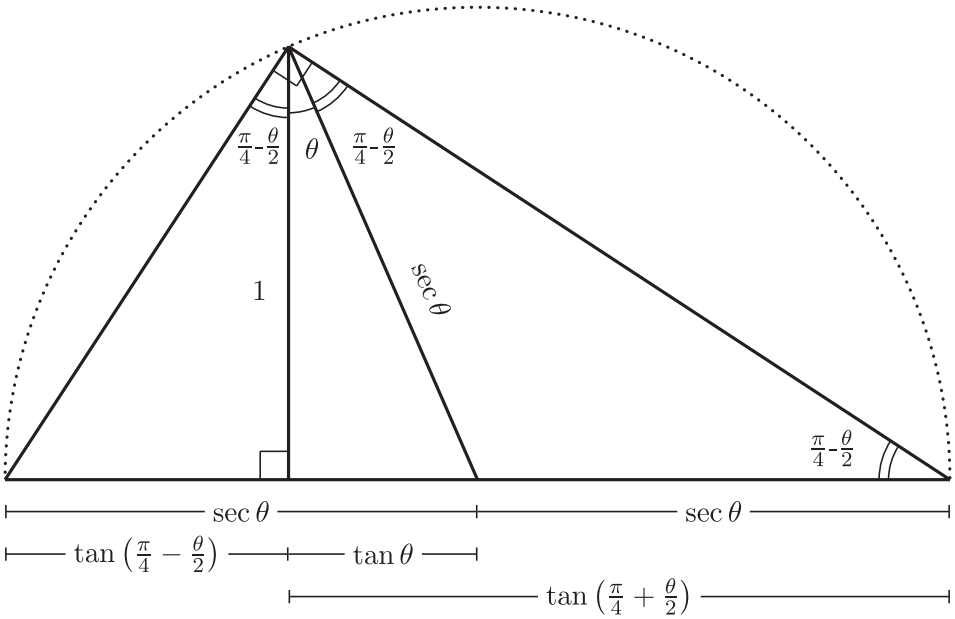


Figure 2 This figure is constructed by merging the two figures in Figure 2, which proves Identities (1) through (3). If we reflect the big triangle about the radius that is perpendicular to the diameter of the semicircle, Identity (4) becomes obvious.

Although we adopted Fincke's version of Identity (1), Bressieu [**3**] had published essentially the same identity two years prior to Fincke (in modern notations, $\tan\theta = \tan 2(\theta-\pi/4)+\sec 2(\theta-\pi/4)$). Romanus utilized Identity (1) to expose errors in Rheticus' tangent and secant tables for values close to 90° [**2**]. All of them used Figure 1 (left) to prove the identity. Identity (2), often appearing together with Identity (1), is proven with Figure 1 (right) [**3**,**5**].

Mathematicians realized the construction of the tangent and secant tables can be achieved by calculating the ratio of sine to cosine (known as *sinum complementi* and several other names at the time) and the inverse of cosine, respectively. However, it is cumbersome to do division. Identities (3) and (4), obtained by adding and subtracting Identities (1) and (2), respectively, solve this problem conveniently. Given the tangents up to 45°, all the tangents and secants can be calculated by simple addition and subtraction using these two identities [**10**]. Briggs, who worked with Napier on the logarithm tables, used Figure 2 (with the semicircle oriented as in Figure 1) to prove Identities (1) through (4) [**4**].

Professor Barry provided a proof to an arctangent identity using Figure 1 (left) [1]. Figure 2 is used to prove several double/half angle identities [7, 11]. Figure 2 can be relabelled to prove the following:

$$\tan\left(\frac{\pi}{4}+\frac{\theta}{2}\right)=\frac{1+\sin\theta}{\cos\theta}=\frac{\cos\theta}{1-\sin\theta} \tag{5}$$

$$\tan\left(\frac{\pi}{4}-\frac{\theta}{2}\right)=\frac{1-\sin\theta}{\cos\theta}=\frac{\cos\theta}{1+\sin\theta} \tag{6}$$

(Hint : Let the radius be 1.)

Acknowledgments The author would like to thank the editor and referees for their suggestions, Dr. Mark Joy for his proofreading, and Ms. Bonnie Ponce for the preparation of the figures.

REFERENCES

[1] Barry, P. (2001). Mathematics without words III. *Coll. Math. J.* 32: 69.
[2] Bockstaele, P. (1992). Adrianus Romanus and the trigonometric tables of Georg Joachinm Rheticus. In: *Amphora: Festschrift für Hans Wussing zu seinem 65. Geburtstag*. Basel/Boston/Berlin: Birkhäuser, pp. 55–66.
[3] Bressieu, M. (1581). Metrices astronomic libri quatuor: Haec maximam partem nova est rerum astronomicarum & geographicarum per plana sphericàque triangula dimensionis ratio, veterique impendio expeditor & compendiosior, apud Aegidium Gorbinum, Parisiis, pp. 39–40.
[4] Briggs, H. (1633). Trigonometria Britannica, sive de doctrina triangulorum libri duo, Excudebat Petrus Rammasenius, Goud, pp. 50–52, 56.
[5] Fincke, T. (1583). Geometriæ rotundi libri XIIII, per Sebastianum Henricpetri, Basileae, pp. 77–78.
[6] Hutton, C. (1811). Mathematical Tables: Containing the Common, Hyperbolic, and Logistic Logarithms; Also Sines, Tangents, Secants, & Versed-sines Both Natural and Logarithmic. Together with Several Other Tables Useful in Mathematical Calculations. To which is Prefixed a Large and Original History of the Discoveries and Writings Relating to Those Subjects; with the Complete Description and Use of the Tables. London: F. C. and Rivington, pp. 1–20.
[7] Nelsen, R. B. (1989). Double-angle formulas. *Coll. Math. J.* 20(1): 1.
[8] Nelsen, R. B. (2015). Proof without words: a trigonometric identity for sec x + tan x. *Math. Mag.* 88: 151.
[9] Otho, V., Rheticus, G. (1596). Opus palatinum de triangulis, Neostadii in Palatinatu: Excudebat Matthus Harnisius, Heidelberg.
[10] Ritter, F. (1895). François Viète, inventeur de l'algebre moderne, 1540-1603, notice sur sa vie et son oeuvre, au Dépot de la Revue occidentale, Paris, p. 49.
[11] Walker, R. J. (1942). Half-angle formulas, *Amer. Math. Monthly*. 49(5): 325.

Summary. The construction of trigonometric tables used to be a challenging activity in mathematics. This article presents a few figures from the sixteenth and seventeenth centuries to illustrate four trigonometric identities that were used in the construction of the tangent and secant tables at the time. Other applications of these figures are also explored.

REX H. WU (MR Author ID: 1293646) is an internist who received his bachelor's degree in mathematics from New York University. After getting married, establishing his private practice and having two children, he finally has some time to get back to mathematics.

Opt-emoji Ghost, Robert Bosch; digital print, 2017. An optimal TSP tour of 1024 points that were arranged to resemble an emoji rendition of a ghost. The optimal tour was obtained with the Concorde TSP Solver.
See interview on page 305.

Robert Bosch: From Dominos To Traveling Salespeople

ALLISON HENRICH
Seattle University
Seattle, WA 98122
henricha@seattleu.edu

Robert (Bob) Bosch is Professor of Mathematics at Oberlin College and an award-winning writer and artist. He specializes in optimization, the field of study concerned with optimal performance. Since 2001, Bosch has devoted increasing amounts of time and effort into devising and refining methods for using optimization to create pictures, portraits, and sculpture. He operates a website, dominoartwork.com, from which it is possible to download free plans for several of his domino mosaics. Images of his work appear on pages 251, 268, 295, and 304.

Q: *What initially inspired you to use your mathematical expertise to create art?*

RB: I was exposed at various points in time to the artwork of Ken Knowlton. The first time was when I was a senior in high school. I opened the May issue of Omni magazine in 1981, and I came across Ken Knowlton's portrait of a domino player. It was made out of 24 complete sets of dominos! I thought this was really cool. Fast forward four years, I'm a senior at Oberlin, hanging out in the math library, and I came across several other domino portraits of Ken Knowlton's. Again, I thought, "Wow! These are really amazing!" Fast forward about 15 years. I'm a professor now at Oberlin College. I'm browsing the stacks at Mudd Library, and I'm looking for cool things to share with my optimization students. I want to convince them that optimization is broadly applicable. So, I come across a Martin Gardner tribute book called "The Mathemagician and the Pied Puzzler." The back cover has a Ken Knowlton domino mosaic portrait of Martin Gardner made out of six complete sets of dominos. At that point in time, I thought to myself, "Oh! I can see how to do that now!" With my training in mathematical optimization, I could see how to make my own domino mosaics. It wouldn't be the same way that Knowlton did it, but there was a way.

Q: *How did you know that the way you were thinking of applying your training was a different method from Knowlton's?*

RB: I didn't know that right at the start, but when I came up with my own method, I started to delve into how Knowlton does it. He patented his method, so I was able to look at his patent application, and it's essentially what we would call a heuristic. It's not guaranteed to produce the best possible arrangement of dominos, where "best" is precisely defined in mathematical terms; it's a relatively simple—and ingenious—procedure that produces really good mosaics. My method is much more computationally intensive—it requires software for solving discrete linear optimization problems. Knowlton's method basically breaks the problem down into two stages. In one stage, he decides how to carve up the portrait into domino-sized regions, not caring which dominos will go where. Knowlton has a clever way of doing that which tries to maximize the total contrast score of his division of the canvas into these domino-sized pieces. The second stage is to actually assign real dominos or "virtual" dominos to those domino-sized pieces. My approach does both stages at once.

Math. Mag. **92** (2019) 305–307. doi:10.1080/0025570X.2019.1624460
MSC: Primary 01A70; Allison Henrich (MR Author ID: 900050)

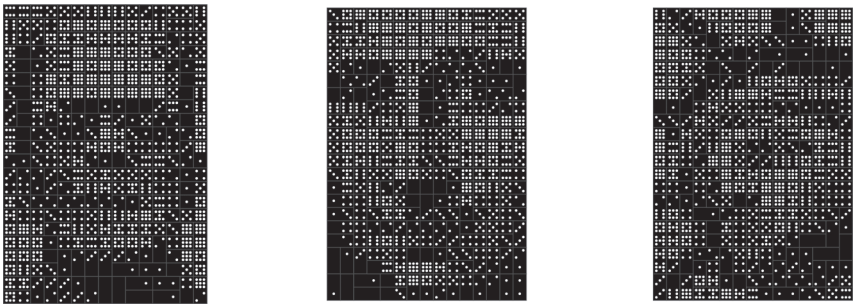


Figure 1 *Three domino mosaics (2014), each made out of three complete sets of (virtual) double-nine dominos: (left) Frankenstein's monster, (center) Leonardo's Mona Lisa, and (right) the Statue of Liberty.*

Q: *You're also really well-known for your art involving traveling salesman problems (TSP). Can you explain a little bit about the mathematics of creating TSP art?*

RB: Sure! You begin with a greyscale image and use an algorithm to convert that image into a point set with some desired number of points. Maybe 1,000 points. Maybe 10,000. Artists call this conversion of an image into a collection of points "stippling." The next step is to imagine that each of these points stands for the location of a city. There's a salesperson who lives in one of these cities. They must visit each of the other cities once and only once and then return home—this is called a tour. Now, we're going to imagine that the salesperson doesn't want to spend too much time on the road. Their goal is to find a tour that travels the least distance or is the least costly. If we solve this person's TSP for this collection of points—cities that are based on the image that we started with—when we draw the tour what we'll get is a continuous line drawing that resembles the initial image. This drawing will actually be a piecewise-linear simple closed curve. If it's close to optimal or optimal, the tour won't cross itself.

Q: *What is an example of a more intricate image that you have represented using this TSP process?*

RB: One of the more satisfying collections of images that I did was a commissioned collection. Professor William Cook is known as one of the world's foremost experts at solving traveling salesman problems to optimality using linear programming-based methods. He's a TSP guru. Back in 2009, he said, "Bob, would you be willing to create a collection of TSP art instances, each with a large number of points? I'd like to put them on my website as a challenge for TSP researchers." Each one of these would be a world record-breaking achievement if someone were to solve it and prove that they had the optimal solution. The smallest of these six problems was based on a segment of Leonardo da Vinci's *Mona Lisa.* It has 100,000 points. The largest is based on Johannes Vermeer's *Girl with a Pearl Earring* and has 200,000 points. There's also a van Gogh, a Botticelli, a Velázquez, and a Courbet. TSP researchers have worked on these problems, and they've produced really good solutions, but so far, nobody has been able to prove that they have the optimal tour for any one of these.

Q: *Are there works of yours that you feel are very "Bob Bosch"—they're just perfectly representative of the type of work that you do?*

RB: Yeah! Two pieces come to mind. I'd say about half of my work is figurative. It starts with some source image and tries to replicate or approximate it in an interesting way. One good example is a piece that I did based on a 1968 photo of me and my dad who, sadly, died the following year. The original image is about two inches wide, but the final TSP art version of it is about four feet wide and three feet high. If you

step way back from this, if you go into the other room and look at it from about 40–50 feet away, it looks remarkably like the original photo. But when you get up close, you realize that it's just this piecewise-linear simple closed curve. It's essentially a distorted circle. Why does this mean so much to me? Well, the content is meaningful. Also, many cultures view circles as symbolizing connectedness, eternity. So, when I view this piece that I made, I can't help but think of it asserting that I am still connected to my father. I have an emotional reaction to it. I think it's a piece that actually succeeds as visual art beyond something that is mathematical.

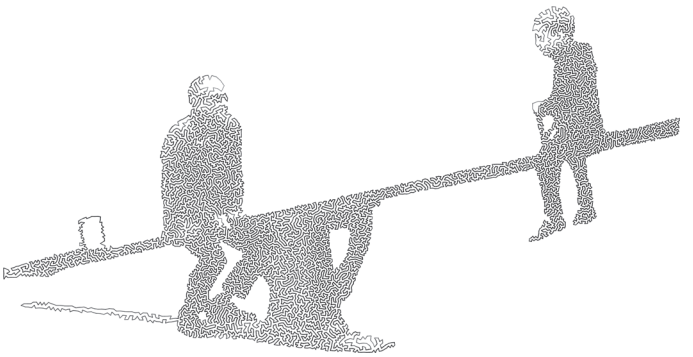


Figure 2 *Father and Son (2013), a continuous line drawing derived from a high-quality solution to a 10,000-point instance of the TSP. Based on a 1968 photo of Robert K. Bosch and Bob Bosch taken by Charlotte Woebcke Bosch.*

Another piece that I like a lot is more in the visual design category. It's called *Embrace.* It's a piece of TSP art. I started with an image that had 6-fold rotational symmetry. It's like a curvy Star of David. Actually, it looks kind of like two fidget spinners that spin together in some unnatural way. I modified both the stippling process and the TSP process because I wanted my point set and my tour to have rotational symmetry, like the original image. Once I had my symmetric simple closed curve, I had the piece cut with a water jet cutter so that the "inside" of the curve is stainless steel and the "outside" is brass. I've given a talk where I'm describing the Jordan curve theorem, and I'll say, "Oh! The Jordan curve theorem says that any simple closed curve divides the plane into two pieces—an inside and an outside." I'm looking at people in the audience, and they're nodding, "Yes. Yes, of course." But then I'll bring out *Embrace*, I'll lift off the outside, and I'll hear: "Ooooh!" It gives a way of experiencing the Jordan curve theorem in a different way. You can actually feel it, experience it kinesthetically. (An image of *Embrace* appears on page X.)

Q: *What are you currently working on?*

RB: I'm just finishing up a book that tries to put into one package how to use mathematical optimization to make visual artwork. It's being published by Princeton University Press in the fall of 2019, and the title is "Opt Art: From Mathematical Optimization to Visual Design." The book is a gentle introduction to linear optimization. It presents some of the classical applications, one of which is the TSP. It also spends a lot of time talking about how I modified or made use of these classical applications and optimization techniques to design visual artwork. I tried to write it so that it could be enjoyed by anyone who's been exposed—at least a little bit—to a first semester course in calculus. I'm imagining that someone who took Differential Calculus and forgot all of it would be able to read my book and get a lot out of it. I'm very excited about it.

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.

Klein, Grady, and Yoram Bauman, *The Cartoon Guide to Calculus*, Hill and Wang, New York, 2019; 207 pp, $18.95. ISBN 978-0-8090-3369-0.

This light introduction to calculus, written by an economist ("the world's first and only stand-up economist") and illustrated by a talented artist, costs about 10% of your average thousand-page calculus textbook. But it's fun to read, focuses on the big ideas, and explains how calculus relates to the rest of mathematics and to economics and physics. It emphasizes the two forms of the fundamental theorem of calculus and ends with a chapter that stresses that mathematics is about finding patterns and "formalizing those patterns with proofs." Students learning how to use and apply calculus will still need to work through the exercises of a textbook, but this book contains some of what you want students to remember 20 years later.

Strogatz, Steven, *Infinite Powers: How Calculus Reveals the Secrets of the Universe*, Houghton Mifflin Harcourt, 2019; xxvi + 360 pp, $28, $16.99(P), $15.99(Kindle). ISBN 978-1-328-879981.

Strogatz, Steven, Outsmarting a virus with math: How calculus helped to drive the fight against HIV, *Scientific American* (April 2019) 70–73.

This book too costs about 10% of a calculus textbook, and it contains a lot more of what you want students to remember 20 years later: "Working behind the scenes, calculus is an unsung hero of modern life." Author Strogatz invents the name "Infinity Principle" to describe calculus as taking the divide-and-conquer strategy to the ultimate degree: namely, calculus makes a hard problem simpler by dividing it into an *infinity* of simpler parts, analyzing those, then reassembling the *infinity* of results back together into a solution to the original problem. Strogatz strolls through the history of calculus, pointing out resulting contemporary applications such as lasers, GPS, computer animation, data compression, predicting outcomes of facial surgery, detecting blood flows in the body, locating tumors, designing airplane wings, and much more. Particularly notable is the role that calculus played in determining that a single drug or even two drugs cannot be effective against HIV, but a three-drug combination can succeed. The book concludes with three examples of the "eerie effectiveness" of calculus in advanced contemporary physics, and ends with the unanswered question "Why is the universe comprehensible, and why is calculus in sync with it?" (The *Scientific American* article, excerpted from the book, is remarkable for the presence of half a dozen equations, which that periodical normally eschews.)

Roberts, David Lindsay, *Republic of Numbers: Unexpected Stories of Mathematical Americans through History*, Johns Hopkins University Press, 2019; viii + 234 pp, $29.95. ISBN 978-1-4214-3308-0.

This book recounts life stories of 23 Americans who were involved in some way with the development and application of mathematics in the U.S. in the past 200 years. A few may be recognizable to this *Magazine*'s readers—Lincoln, Gibbs, Ladd-Franklin, Hollerith, Hopper, Nash—but the others too played roles that author Roberts describes.

Math. Mag. **92** (2019) 308–309. doi:10.1080/0025570X.2019.1648112

Dunbar, Steven R., *Mathematical Modeling in Economics and Finance: Probability, Stochastic Processes, and Differential Equations*, MAA Press, 2019; xv + 223 pp, $75 ($56.25 to MAA members). ISBN 978-1-4704-4839-4.

This book provides "a textbook for a capstone course focusing on mathematical modeling in economics and finance." Although the book does not require any previous knowledge about financial instruments, author Dunbar makes clear that the reader should have background in mathematical probability, statistical estimation, and basic theory of interest (compounding, present values, annuities). Topics include binomial models, gambler's ruin, Brownian motion, stochastic calculus, and the Black–Scholes equation. Simulation scripts in R are included, and there are exercises (without answers). Dunbar advises that the book "intentionally addresses" the cognitive and content recommendations in the MAA CUPM (Committee on the Undergraduate Program in Mathematics) guidelines, which he quotes at length.

Huckle, Thomas, and Tobias Neckel, *Bits and Bugs: A Scientific and Historical Review of Software Failures in Computational Science*, SIAM, 2019; xii + 251 pp, $44. ISBN 978-1-611975-55-0.

Computation is part of mathematics. This book examines in fascinating detail many notable (and very regrettable) failures in scientific computing: crashes of planes, rockets, and space vehicles; failures of the Patriot missile system; the London Millennium Bridge wobble; the 2008 financial crisis; accidents with self-driving cars; the Pentium computer chip; medical radiation overdoses; the Denver Airport baggage-handling fiasco. The authors describe each failure in detail, trace its cause, and cite numerous failures of a similar nature. They provide in addition simple examples of the phenomenon, Matlab code to reproduce similar behavior, and QR codes that link to videos and animations.

Hodge, Jonathan K., and Richard E. Klima, *The Mathematics of Voting and Elections: A Hands-On Approach*, 2nd ed., American Mathematical Society, 2018; xiii + 238 pp, $52(P) ($46.80 to MAA members). ISBN 978-1-4704-4287-3.

It seems that elections are always in season in the U.S.! This book offers mathematical perspectives on that fundamental instrument of democracy. Several chapters on mathematical theory of voting lead up to Arrow's impossibility theorem. One chapter new to this second edition treats new developments in mathematical analysis of gerrymandering, and another delves into strategic voting and other manipulation of voting systems, leading to a corollary of Arrow's theorem. Other chapters treat power indexes, the U.S. Electoral College, the history of U.S. apportionment methods, and hazards in referendum elections. The authors wanted to make the text accessible to a non-mathematical audience, so they include no worked-out examples or exercises but instead provide questions aimed at ensuring understanding of concepts (with answers), and other questions to further critical thinking and analysis (no answers), often asking the reader to do research on a particular election and write a summary of findings.

Harris, Michael, Why the proof of Fermat's last theorem doesn't need to be enhanced, https://www.quantamagazine.org/why-the-proof-of-fermats-last-theorem-doesnt-need-to-be-enhanced-20190603.

It is now 26 years since Andrew Wiles announced proof of Fermat's last theorem. The proof used concepts that are esoteric even to most mathematicians. Should you believe that the result is established? Would it be worthwhile to apply proof-verification software to his proof, as Thomas Hales accomplished in 2014 for his 1998 proof of the Kepler conjecture about dense packing of spheres? This article gives a very clear overall outline of the strategy and major ingredients in the Wiles proof. Author Harris also relates that, perhaps curiously, number theorists appear to show no interest in the application of proof-verification software to the proof. Rather, they regard the proof as a "point of departure for an open-ended dialogue that is too elusive and alive to be limited by foundational constraints that are alien to the subject matter."

PROBLEMS

EDUARDO DUEÑEZ, *Editor*
University of Texas at San Antonio

EUGEN J. IONAŞCU, *Proposals Editor*
Columbus State University

JOSÉ A. GÓMEZ, Facultad de Ciencias, UNAM, Mexico; *CODY PATTERSON*, Texas State University; *RICARDO A. SÁENZ*, Universidad de Colima, Mexico; *ROGELIO VALDEZ*, Centro de Investigación en Ciencias, UAEM, Mexico; *Assistant Editors*

Problem 2067 Updated *Editor's Note.* The statement of Problem 2067 that appeared in the April 2019 issue omitted the critical hypothesis that chord $\overline{MN}$ goes through P. We sincerely regret the mistake, and thank Robert Calcaterra for bringing it to our attention. The corrected statement of Problem 2067 appears below.

2067. *Proposed by Elton Bojaxhiu, Eppstein am Taunus, Germany and Enkel Hysnelaj, Sydney, Australia.*

Chord $\overline{XY}$ of a circle $\mathcal{C}$ is not a diameter. Let P, Q be two different points strictly inside $\overline{XY}$ such that Q lies between P and X. Chord $\overline{MN}$ through P is perpendicular to the diameter of $\mathcal{C}$ through Q, where $MP < NP$. Prove that $(MQ - PQ) \cdot XY \geq 2 \cdot QX \cdot PY$, and characterize those cases in which equality holds.

Proposals

To be considered for publication, solutions should be received by March 1, 2020.

2076. *Proposed by Michael Goldenberg, The Ingenuity Project, Baltimore Polytechnic Institute, Baltimore, MD and Mark Kaplan, Towson University, Towson, MD.*

Given real numbers C_0, C_1, and C_2, one defines a *general Tribonacci (GT) sequence* $\{C_n\}$ recursively by the relation $C_{n+3} = C_{n+2} + C_{n+1} + C_n$ for all $n \geq 0$. Such GT-sequence $\{C_n\}$ is *nonsingular* if

$$\Delta = \begin{vmatrix} C_0 & C_1 & C_2 \\ C_1 & C_2 & C_3 \\ C_2 & C_3 & C_4 \end{vmatrix} \neq 0.$$

Math. Mag. **92** (2019) 310–317. doi:10.1080/0025570X.2019.1648111

We invite readers to submit original problems appealing to students and teachers of advanced undergraduate mathematics. Proposals must always be accompanied by a solution and any relevant bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. Submitted problems should not be under consideration for publication elsewhere.

Proposals and solutions should be written in a style appropriate for this MAGAZINE.

Authors of proposals and solutions should send their contributions using the Magazine's submissions system hosted at http://mathematicsmagazine.submittable.com. *More detailed instructions are available there. We encourage submissions in PDF format, ideally accompanied by LaTeX source. General inquiries to the editors should be sent to* mathmagproblems@maa.org.

A *dual Tribonacci (DT) sequence* $\{D_n\}$ is one that satisfies the dual recurrence $D_{n+3} + D_{n+2} + D_{n+1} = D_n$ for $n \geq 0$. Show that for any nonsingular GT-sequence $\{C_n\}$ with C_0, C_1, C_2 positive there exists a DT-sequence $\{D_n\}$ such that, for all $n \geq 0$,

$$\arctan\left(\frac{\sqrt{D_n}}{C_n}\right) = \arctan\left(\frac{\sqrt{D_n}}{C_{n+1}}\right) + \arctan\left(\frac{\sqrt{D_n}}{C_{n+2}}\right) + \arctan\left(\frac{\sqrt{D_n}}{C_{n+3}}\right).$$

2077. *Proposed by Li Zhou, Polk State College, Winter Haven, FL.*

Prove that in any triangle with side lengths a, b, c, inradius r, and circumradius R, we have

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} + \frac{r}{R} > \frac{5}{3}.$$

2078. *Proposed by Florin Stanescu, Serban Cioculescu School, Gaesti, Romania.*

Let A, B be $n \times n$ complex matrices such that $A^2 + B^2 = 2AB$. Prove that $(AB - BA)^m = \mathbf{0}$ for some $m \leq \left\lceil \frac{n}{2} \right\rceil$.

2079. *Proposed by Ovidiu Furdui and Alina Sîntămărian, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.*

Given real numbers a, b, with $b > 0$, prove that the integral

$$J(a,b) := \int_0^\infty \left[2 + (x+a)\ln\left(\frac{x}{x+b}\right)\right] dx$$

converges if and only if $a = 1$ and $b = 2$, and find the value $J(1,2)$.

2080. *Proposed by the UTSA Problem Solving Club, University of Texas at San Antonio, San Antonio, TX.*

For $n \geq 3$, let W_n be the wheel graph consisting of an n-cycle all whose vertices are joined to an additional distinct vertex.

(*i*) How many colorings of the $2n$ edges of W_n using $k \geq 2$ colors result in no monochromatic triangles?

(*ii*) Regard two colorings of W_n as equivalent if there is a graph automorphism of W_n that maps the first coloring to the second. If $k \geq 2$ and $p > 3$ is prime, count all non-equivalent colorings of W_p using k colors.

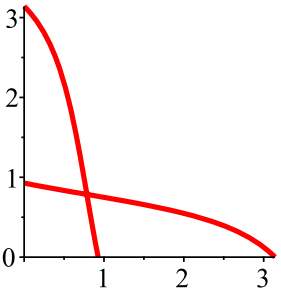
Quickies

1093. *Proposed by Mihaly Bencze, Brasov, Romania.*

Show that 2019^{2n} can be expressed as a sum of ten different positive squares, for every positive integer n.

1094. *Proposed by Julien Sorel, Piatra Neamt, PNI, Romania.*

The curve $2\sin(x+y) - \cos(x-y) = 1$ has a self-intersection point at $(\pi/4, \pi/4)$ as shown in the figure below. Find the angle between the two tangent lines to the curve at this point.



Solutions

The largest roots of a sequence of polynomials **October 2018**

2051. *Proposed by Ángel Plaza, Universidad de Las Palmas de Gran Canaria, Spain.*

For any positive integer n consider the polynomial $P_n(x) = x^4 - nx^3 - nx^2 - nx + 1$ and let a_n be the largest of its real roots. Find

$$\lim_{n\to\infty} \frac{a_1 + a_2 + \cdots + a_n}{n^2}.$$

Solution by José Heber Nieto, Universidad del Zulia, Maracaibo, Venezuela.
We show that the limit exists and equals $1/2$. If $x \geq n + 1$, then

$$\begin{aligned} P_n(x) &= (x-n)x^3 - nx^2 - nx + 1 \geq 1x^3 - nx^2 - nx + 1 \\ &= (x-n)x^2 - nx + 1 \geq 1x^2 - nx + 1 = (x-n)x + 1 \\ &\geq 1x + 1 \geq n + 2 > 0. \end{aligned}$$

On the other hand, $P_n(n) = -n^3 - n^2 + 1 < 0$. By continuity of P_n and the intermediate value theorem, it follows that $n < a_n < n + 1$; hence,

$$\frac{n^2+n}{2n^2} = \frac{1}{n^2}\sum_{i=1}^{n} i < \frac{1}{n^2}\sum_{i=1}^{n} a_i < \frac{1}{n^2}\sum_{i=1}^{n}(i+1) = \frac{n^2+3n}{2n^2}.$$

The first and last expressions above have the same limit $1/2$ as n tends to infinity. By the sandwich theorem,

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} a_i = \frac{1}{2}.$$

Also solved by Ulrich Abel (Germany), Terrance Alvarez & Cyane Gonzalez, Michael A. Ask, Michel Bataille (France), Necdet Batir (Turkey), Brian D. Beasley, Anthony J. Bevelacqua, Robert Calcaterra, Robin Chapman (UK), Jyoti Champanerkar, John Christopher, Michael P. Cohen, Bill Cowieson, Antonella Cupillari, Richard Daquila, Robert L. Doucette, Dmitry Fleischman, Charles Fleming, Natacha Fontes-Merz, Michael Goldenberg & Mark Kaplan, Abhay Goel, Dean Gooch, Lixing Han, Kyle Hansen, GWstat Problem Solving Group, Eugene A. Herman, Theo Koupelis, Elias Lampakis (Greece), Jeffery M. Lewis, James Magliano, Peter McPolin (Northern Ireland), Northwestern University Math Problem Solving Group, Michael Reid, Volkhard Schindler, Joel Schlosberg, Edward Schmeichel, Mark Schultz, Randy K. Schwartz, Achilleas Sinefakopoulos (Greece), Nicholas C. Singer, Albert Stadler (Switzerland), David Stone & John Hawkins, Koopa

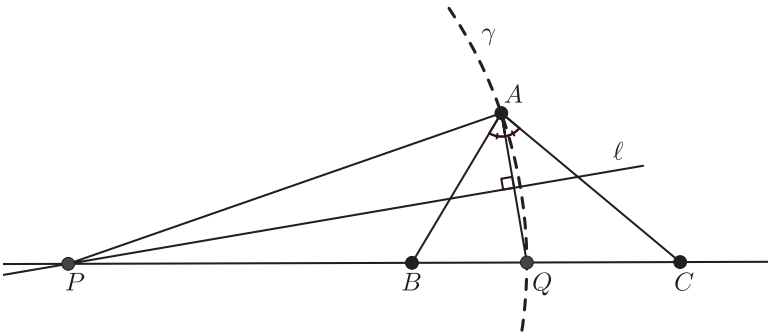
Tak Lun Koo (Hong Kong), The Iowa State Undergraduate Problem Solving Group, Michael Vowe (Switzerland), John Zacharias, and the proposer. There was one incomplete or incorrect solution.

A pencil of lines obtained from any scalene triangle **October 2018**

2052. *Proposed by Michel Bataille, Rouen, France.*

Let $\triangle ABC$ be a scalene triangle. Let D be a variable point on line $\overleftrightarrow{BC}$ such that $D \neq B$ and $D \neq C$. Let E lie on $\overleftrightarrow{BC}$ so $\overleftrightarrow{AE}$ is the reflection of $\overleftrightarrow{AD}$ across the bisector of angle $\angle BAC$. Let O_1, O_2 be the circumcenters of triangles $\triangle ABD$ and $\triangle ACE$, respectively. Prove that there exists a point P, independent of the choice of D, such that line $\overleftrightarrow{O_1O_2}$ passes through P.

Solution by Peter McPolin, St. Mary's University College, Northern Ireland.
Let the bisector of angle $\angle BAC$ intersect $\overline{BC}$ at Q. Since $\triangle ABC$ is not isosceles, the perpendicular bisector ℓ of the segment $\overline{AQ}$ is not parallel to $\overleftrightarrow{BC}$. The point P of intersection of ℓ and $\overleftrightarrow{BC}$ depends only on the choice of the triangle $\triangle ABC$. We show that O_1, O_2, and P are collinear. Let γ be the circle with centre P passing through A (hence also through Q). We prove that the points B and C are inverses with respect to γ.



In the figure above (where $AB < AC$, which may be assumed to hold without loss of generality), we have $\angle PAB + \angle BAQ = \angle PAQ = \angle AQP$ ($\triangle PAQ$ is isosceles with $\overline{PA} = \overline{PQ}$), $\angle BAQ = \angle QAC$ ($\overline{AQ}$ is the bisector of angle $\angle BAC$), $\angle AQP = \angle QAC + \angle ACQ$ ($\angle AQP$ is an exterior angle of triangle $\triangle QAC$), and so $\angle PAB = \angle ACQ = \angle ACP$. Thus, triangles $\triangle PAB$ and $\triangle PCA$ are similar; hence, $PB/PA = PA/PC$, so $PB \cdot PC = PA^2$, showing that B and C are inverses with respect to γ.

If D lies on the half-line $\overrightarrow{PQ}$ then, by construction of E from D, the same line $\overleftrightarrow{AQ}$ bisects angle $\angle EAD$, so the argument above proves that D and E are inverses with respect to γ. If D lies on the other half-line $\overrightarrow{PQ'}$, where Q' is the diametrical opposite of Q on γ, the same conclusion follows upon replacing Q by Q' in the preceding argument. (If $D = P$ then the reflection of $\overleftrightarrow{AD}$ on $\overleftrightarrow{AQ}$ is parallel to $\overleftrightarrow{BC}$, so E is undefined—it may be conventionally regarded as the point at infinity, inverse with respect to γ of its center P.) Inversion with respect to γ fixes A, so this inversion transforms the circumcircle δ of triangle $\triangle ABD$ into the circumcircle ε of triangle $\triangle ACE$. Evidently, the circles $\gamma, \delta, \varepsilon$ are coaxial, so their centers P, O_1, and O_2 are collinear; moreover, as indicated above, P is independent of the choice of D.

Also solved by Andrea Fanchini (Italy), Elias Lampakis (Greece), José H. Nieto (Venezuela), Achilleas Sinefakopoulos (Greece), Elton Bojaxhiu (Albania) & Enkel Hysnelaj (Australia) Peter

McPolin (Northern Ireland), Lienhard Wimmer (Germany), Theo Koupelis, Kyle Gatesman and the proposer. There was one incomplete or incorrect solution.

Maximally deranged permutations **October 2018**

2053. *Proposed by Sung Soo Kim, Hanyang University, Korea.*

Let $a = (a_1, a_2, \ldots, a_{2018})$ be a permutation of the integers $1, 2, \ldots, 2018$. For any integer k in the range $1 \le k \le 2018$, let $l_k(a)$ be the length of the longest monotone subsequence of $(a_k, a_{k+1}, \ldots, a_{2018})$ whose first term is a_k, and let $L(a) = \sum_{k=1}^{2018} l_k(a)$. Find the minimum value of $L(a)$ as a ranges over all permutations of $1, 2, \ldots, 2018$.

Solution by Michael Reid, University of Central Florida, Orlando, FL.

The minimum value is $\sum_{k=1}^{2018} \left\lceil \sqrt{k} \right\rceil = 61\,440$. We need the following well-known result from combinatorics.

Theorem [P. Erdős, G. Szekeres, A combinatorial problem in geometry, *Compositio Mathematica,* **2** (1935) 463–470, https://eudml.org/doc/88611]

Let r, s be natural numbers. A sequence of distinct real numbers having length $> rs$ has either an increasing subsequence of length $> r$, or a decreasing subsequence of length $> s$.

Proof. For a sequence $a = (a_1, a_2, \ldots, a_n)$ with $n > rs$, define $f, g : \{1, \ldots, n\} \to \mathbb{N}$ as follows: $f(i)$ (resp., $g(i)$) is the length of the longest increasing (resp., decreasing) subsequence of $(a_i, \ldots, a_n)$ with first term a_i. The pairs $(f(i), g(i))$ as i varies are all distinct: If $i < j$ and $a_i < a_j$ (resp., $a_i > a_j$), then $f(i) > f(j)$ (resp., $g(i) > g(j)$). By the pigeonhole principle, since $n > rs$, the pairs $(f(i), g(i))$ cannot all lie in $\{1, \ldots, r\} \times \{1, 2, \ldots, s\}$; thus, either $f(i) > r$ or $g(i) > s$ for some i, whence the conclusion of the theorem follows immediately.

Resuming the solution, for $n \in \mathbb{N}$ and any sequence $a = (a_1, a_2, \ldots, a_n)$ of distinct real numbers, define $l_k(a)$ as the length of the longest monotone subsequence of $(a_k, a_{k+1}, \ldots, a_n)$ whose first term is a_k, and $L(a)$ as $\sum_{k=1}^{n} l_k(a)$. By induction on n, we will show that

$$L(a) \ge M(n) := \sum_{k=1}^{n} \left\lceil \sqrt{k} \right\rceil \qquad (*)$$

for every such sequence a of length n. For $n = 1$, inequality $(*)$ holds since both its sides are equal to 1. Next, suppose inequality $(*)$ holds for all sequences of some fixed length n, and let a be a sequence of length $n + 1$. For $r = s = \left\lceil \sqrt{n+1} \right\rceil - 1$, the sequence a has length $n + 1 \ge rs + 1 > r^2$, so its longest monotone subsequence has length (at least) $r + 1 = \left\lceil \sqrt{n+1} \right\rceil$. Let a_i be the first term of a longest such subsequence, so $l_i(a) \ge \left\lceil \sqrt{n+1} \right\rceil$, and let $\hat{a} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{n+1})$ be the length-n sequence obtained from a by deleting the term a_i. For $i < k \le n + 1$, we have $l_k(a) = l_{k-1}(\hat{a})$ (= length of the longest monotone subsequence of $(a_k, a_{k+1}, \ldots, a_{n+1})$ whose first term is a_k). For $1 \le k < i$, we have $l_k(a) \ge l_k(\hat{a})$ because any monotone subsequence of $\hat{a}$ starting at a_k is also a monotone subsequence of a starting at a_k.

It follows that

$$L(a) = l_i(a) + \sum_{k=1}^{i-1} l_k(a) + \sum_{k=i+1}^{n+1} l_k(a) \geq \left\lceil \sqrt{n+1} \right\rceil + \sum_{k=1}^{i-1} l_k(\hat{a}) + \sum_{k=i+1}^{n+1} l_{k-1}(\hat{a})$$

$$= \left\lceil \sqrt{n+1} \right\rceil + \sum_{k=1}^{n} l_k(\hat{a}) = \left\lceil \sqrt{n+1} \right\rceil + L(\hat{a})$$

$$\geq \left\lceil \sqrt{n+1} \right\rceil + M(n) = M(n+1),$$

by the assumed validity of $(*)$ for the length-n sequence $\hat{a}$. This completes the inductive proof of $(*)$ for all $n \geq 1$.

Call a sequence $a = (a_1, a_2, \ldots, a_n)$ of n distinct numbers *deranged* if $l_k(a) \leq \left\lceil \sqrt{n+1-k} \right\rceil$ for $1 \leq k \leq n$. A deranged sequence satisfies the inequality $L(a) \leq \sum_{k=1}^{n} \left\lceil \sqrt{n+1-k} \right\rceil = M(n)$. By inequality $(*)$, a deranged sequence actually satisfies that $L(a) = M(n)$ is minimum among all sequences of length n.

First, we construct deranged sequences whose length n is an arbitrary perfect square. Any sequence of length $1^2 = 1$ is deranged. Assume a deranged sequence a of length $n = t^2$ has been constructed; we proceed to construct a deranged sequence $\hat{a}$ of length $N = (t+1)^2$. For any choice of $b_1, b_2, \ldots, b_t, b_{t+1}$ and $c_1, c_2, \ldots, c_t$ such that $\min\{a_1, \ldots, a_n\} > b_1 > b_2 > \cdots > b_{t+1}$ and $\max\{a_1, \ldots, a_n\} < c_1 < c_2 < \cdots < c_t$, construct the sequence

$$\hat{a} = (b_1, b_2, \ldots, b_{t+1}, c_1, c_2, \ldots, c_t, a_1, a_2, \ldots, a_n),$$

which we proceed to show is deranged. The sequence $\hat{a}$ has length $(t+1) + t + t^2 = (t+1)^2 = N$. Consider a monotone subsequence of $\hat{a}$ starting at some $b_i = \hat{a}_i$. If the subsequence contains a second term b_j, then it is necessarily decreasing, and thus a subsequence of $(b_1, b_2, \ldots, b_{t+1})$ (since each b_j is less than every a_k and every c_i by construction) and hence has length at most $t+1$. If the subsequence does not contain a second term b_j, but contains a term c_j, then it is necessarily increasing, so it is a subsequence of $(b_i, c_1, c_2, \ldots, c_t)$ (since each c_j is greater than every a_k by construction), and thus has length at most $t+1$. If the subsequence does not contain a second term b_j, nor any term c_j, then it consists of b_i followed by a decreasing subsequence of a; such a subsequence starting with b_i has length at most $1 + \max\{l_1(a), \ldots, l_n(a)\} \leq 1 + t$. Hence, $l_k(\hat{a}) \leq t + 1 = \left\lceil \sqrt{N+1-k} \right\rceil$ for $1 \leq k \leq t+1$. Similar consideration of a monotonic subsequence starting with some $c_i = \hat{a}_{t+1+i}$ shows that $l_k(\hat{a}) \leq \left\lceil \sqrt{N+1-k} \right\rceil$ for $t+2 \leq k \leq 2t+1$. For $2t+1 < k \leq n$, we have $l_k(\hat{a}) = l_{k-(2t+1)}(a) \leq \left\lceil \sqrt{n+1-(k-(2t+1))} \right\rceil = \left\lceil \sqrt{N+1-k} \right\rceil$ since a is deranged by hypothesis, hence $\hat{a}$ is a deranged sequence of length $N = (t+1)^2$.

To obtain a deranged sequence a of arbitrary (non-square) length n, it suffices to take the last n terms of a deranged sequence of length $t^2 \geq n$. Finally, to obtain a deranged permutation of $\{1, 2, \ldots, n\}$, let a be a length-n deranged sequence and let $\sigma \in S_n$ be the "sorting" permutation of a, so $a_{\sigma(1)} < a_{\sigma(2)} < \cdots < a_{\sigma(n)}$. The sequence $\sigma^{-1} = (\sigma^{-1}(1), \sigma^{-1}(2), \ldots, \sigma^{-1}(n))$ has the same relative ordering as the sequence $(a_1, a_2, \ldots, a_n)$, and thus σ^{-1} is a deranged permutation of $\{1, 2, \ldots, n\}$. To conclude the solution, let a be a deranged permutation of $\{1, 2, \ldots, 2018\}$. Then, $L(a) = M(2018) = 61\,440$ is minimum among all permutations.

Also solved by José Nieto (Venezuela), and the proposer. There were 2 incomplete or incorrect solutions.

A second-moment inequality when first moment is zero **October 2018**

2054. *Proposed by Florin Stanescu, Şerban Cioiculescu school, Găeşti, Romania.*

Let $f : [0, 1] \to \mathbb{R}$ be differentiable with bounded derivative. If $\int_0^1 xf(x)dx = 0$, prove that

$$36 \cdot \left| \int_0^1 x^2 f(x)dx \right| \leq \sup_{x\in[0,1]} |f'(x)|.$$

Solution by Lixing Han, University of Michigan-Flint, Flint, MI.
Integrating by parts, we have

$$\int_0^1 x^2 f'(x)\,dx = x^2 f(x)\big|_0^1 - 2\int_0^1 xf(x)\,dx = f(1),$$

since $\int_0^1 xf(x)\,dx = 0$ by hypothesis. Integrating by parts again:

$$\int_0^1 x^2 f(x)\,dx = \frac{1}{3}\,x^3 f(x)\big|_0^1 - \frac{1}{3}\int_0^1 x^3 f'(x)\,dx = \frac{1}{3}f(1) - \frac{1}{3}\int_0^1 x^3 f'(x)\,dx.$$

Solving for $f(1)$ in this equation and combining with the first above, we obtain

$$\int_0^1 x^2 f(x)\,dx = \frac{1}{3}\int_0^1 x^2 f'(x)\,dx - \frac{1}{3}\int_0^1 x^3 f'(x)\,dx = \frac{1}{3}\int_0^1 (x^2 - x^3) f'(x)\,dx.$$

Therefore,

$$\left|\int_0^1 x^2 f(x)\,dx\right| = \frac{1}{3}\left|\int_0^1 (x^2 - x^3) f'(x)\,dx\right| \leq \frac{1}{3}\int_0^1 (x^2 - x^3)\left|f'(x)\right| dx$$
$$\leq \frac{1}{3}\int_0^1 (x^2 - x^3)\,dx \cdot \sup_{0\leq x\leq 1} |f'(x)| = \frac{1}{36} \cdot \sup_{0\leq x\leq 1} |f'(x)|.$$

The inequality asserted in the statement of the problem follows immediately.

Also solved by Ulrich Abel (Germany), Michel Bataille (France), Robin Chapman (UK), Gary Chung, Michael P. Cohen, Robert Calcaterra, William Cowieson, Souvik Dey, Robert Doucette, Eugene Herman, Elgin Johnston, Koopa Koo (Hong Kong), Elias Lampakis (Greece), Kee-Wai Lau (Hong Kong), Joel Schlosberg, Ioannis Sfikas (Greece), Nicholas Singer, Albert Stadler (Switzerland), Michael Vowe (Switzerland), Scott Wolf, Shazeena Ashraf, Robert Summers, Braeden Duke & Matthew Cullum and the proposer.

Cyclic groups via characteristic subgroups **October 2018**

2055. *Proposed by Ioan Băetu, Botoşani, Romania.*

Let n be a cube-free positive integer. Assume that G is a finite group of order n such that for every subgroup H of G and every automorphism f of H, the equality $K = \{f(x) : x \in K\}$ holds for every subgroup K of H. Prove that G is cyclic.

Solution by Anthony J. Bevelacqua, University of North Dakota, Grand Forks, ND.
Suppose $x, y \in G$ satisfy $\langle x\rangle \cap \langle y\rangle = \{1\}$. By hypothesis, the conjugation automorphism $z \mapsto x^{-1}zx$ of G fixes $\langle y\rangle$, hence $x^{-1}yx \in \langle y\rangle$, and similarly $y^{-1}x^{-1}y \in \langle x^{-1}\rangle = \langle x\rangle$. It follows that $x^{-1}y^{-1}xy \in \langle x\rangle \cap \langle y\rangle = \{1\}$, so x and y commute.

Next, we show that, for any prime p dividing n, a Sylow p-subgroup P of G is cyclic. Denote by C_k the cyclic group of order $k \geq 1$. Since n is cube-free, P has order p or p^2; thus, P is isomorphic to one of the cyclic groups C_p, C_{p^2}, or the non-cyclic group $C_p \times C_p$. The subgroup $C_p \times \{1\}$ of $C_p \times C_p$ is not fixed by the automorphism $(x, y) \mapsto (y, x)$; thus, the hypothesis on G implies that P is not isomorphic to $C_p \times C_p$, so P is cyclic.

To conclude the proof, let $p_1, \ldots, p_r$ be the distinct primes dividing n. For $j = 1, \ldots, r$, let x_j be a generator of a Sylow p_j-subgroup of G. By the first Sylow theorem, we have $|x_1| \cdots |x_r| = n$. The elements $x_1, \ldots, x_r$ have pairwise coprime orders, hence generate groups with pairwise trivial intersection. By the argument in the first paragraph above, these elements commute pairwise, and furthermore $|x_1 \cdots x_r| = |x_1| \cdots |x_r| = n$. Hence, G is cyclic generated by $x_1 \cdots x_r$.

Editor's Note. Michael Reid pointed out that the hypothesis that G is finite may be relaxed to finitely generated (but not to infinitely generated). The conclusion that G is cyclic then follows from a more delicate argument using Baer's theorem.

Also solved by Robert Calcaterra, Robert Doucette, Abhay Goel, Koopa Koo (Hong Kong), José Nieto (Venezuela), Michael Reid, Nikhil Sahoo, Jacob Siehler, and the proposer.

Answers (Solutions to the Quickies from page 311.)

A1093. We have

$$2019^2 = 1480^2 + 969^2 + 555^2 + 485^2 + 455^2 + 300^2 + 200^2 + 185^2 + 150^2 + 100^2.$$

Therefore, for all $n > 0$, letting $m = n - 1 \geq 0$,

$$\begin{aligned}
2019^{2n} &= 2019^{2(m+1)} = 2019^2 \cdot 2019^{2m} \\
&= (1480^2 + 969^2 + 555^2 + 485^2 + 455^2 + 300^2 + 200^2 + 185^2 + 150^2 + 100^2) \cdot 2019^{2m} \\
&= (1480 \cdot 2019^m)^2 + (969 \cdot 2019^m)^2 + (555 \cdot 2019^m)^2 + (485 \cdot 2019^m)^2 \\
&\quad + (455 \cdot 2019^m)^2 + (300 \cdot 2019^m)^2 + (200 \cdot 2019^m)^2 + (185 \cdot 2019^m)^2 \\
&\quad + (150 \cdot 2019^m)^2 + (100 \cdot 2019^m)^2.
\end{aligned}$$

A1094. Implicit differentiation with respect to x gives $2(1 + y') \cos(x + y) + (1 - y') \sin(x - y) = 0$; hence,

$$\nu := \frac{y' + 1}{y' - 1} = \frac{\sin(x - y)}{2 \cos(x + y)}.$$

Using trigonometric identities and the relation $2 \sin(x + y) - \cos(x - y) = 1$, we obtain

$$\begin{aligned}
\nu^2 &= \frac{\sin^2(x - y)}{4 \cos^2(x + y)} = \frac{[1 + \cos(x - y)][1 - \cos(x - y)]}{4[1 + \sin(x + y)][1 - \sin(x + y)]} \\
&= \frac{2 \sin(x + y)}{2[1 + \sin(x + y)]} \cdot \frac{1 - \cos(x - y)}{2 - 2 \sin(x + y)} = \frac{\sin(x + y)}{1 + \sin(x + y)}.
\end{aligned}$$

Thus, at the double point $(\pi/4, \pi/4)$, we have $\nu^2 = \sin(\pi/2)/[1 + \sin(\pi/2)] = 1/2$, so $\nu = \pm 1/\sqrt{2}$. Either of the tangent line slopes $m = y'$ at the double point is related to the respective inclination angle θ by $m = \tan\theta$, while $\nu = (\tan\theta + 1)/(\tan\theta - 1) = \tan(-\theta - \pi/4)$. It follows that the angle sought, equal to the difference of the inclination angles θ_1, θ_2, is equal to $\theta_2 - \theta_1 = (-\pi/4 - \theta_1) - (-\pi/4 - \theta_2) = \arctan(1/\sqrt{2}) - \arctan(-1/\sqrt{2}) = 2 \arctan(1/\sqrt{2}) \approx 70.53°$.

Carl B. Allendoerfer Awards

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of articles of expository excellence published in MATHEMATICS MAGAZINE. The Award is named for Carl B. Allendoerfer, a distinguished mathematician at the University of Washington and president of the Mathematical Association of America, 1959–60.

William Dunham

"The Early (and Peculiar) History of the Möbius Function." MATHEMATICS MAGAZINE, Volume 91, Number 2, April 2018, pages 83–91.

Upon first encountering the Möbius function in a number theory course, a student might well find it "neither useful nor obvious," as the author writes in this lively story of the early appearances of the function. Seeking to perform a kind of inversion on infinite series, August Ferdinand Möbius was led in 1832 to introduce a certain function $\mu(x)$ on the positive integers—although as the author notes, the now-ubiquitous notation $\mu(x)$ only took hold after being introduced some 40 years later by Franz Mertens. In order to achieve his inversions, Möbius found that $\mu(x)$ must satisfy:

1. $\mu(1) = 1$,
2. $\mu(n) = 0$ if n is divisible by the square of some prime, and
3. $\mu(n) = (-1)^r$ if n is a product of r distinct primes.

While it seems non-intuitive at first, Möbius' function gets at something fundamental about the integers—both the prime number theorem and the Riemann hypothesis can be recast as statements about $\mu(x)$. The author leads us on a tour of Möbius' construction from his 1832 paper, and illustrates the power of Möbius' method by showing how the complicated-looking sum

$$\sum_{k=1}^{\infty} \frac{\mu(k)x^k}{1-x^k}$$

equals, remarkably, x.

But the story doesn't end there, and those familiar with the author's previous work might have an inkling where it goes. Some 80 years earlier, Euler had already met the same function. In his typical fashion, Euler derived eye-popping formulas for sums such as $\sum_{k=1}^{\infty} \frac{\mu(k)}{k}$ and $\sum_{k=1}^{\infty} \frac{\mu(k)}{k^2}$ without ever explicitly defining $\mu(x)$. He considered the interplay between infinite products and infinite sums, starting with the product

$$\left(1-\frac{1}{2^n}\right)\left(1-\frac{1}{3^n}\right)\left(1-\frac{1}{5^n}\right)\cdots$$

and its reciprocal. Niftily using the unique factorization of integers into products of prime powers, he showed that this product is given by $\sum_{k=1}^{\infty} \frac{\mu(k)}{k}$ and its reciprocal by the harmonic series. And presto: $\sum_{k=1}^{\infty} \frac{\mu(k)}{k} = 0$. With similar verve, Euler derives

Math. Mag. **92** (2019) 318–320. doi:10.1080/0025570X.2019.1616996
MSC: Primary 01A70

$\sum_{k=1}^{\infty} \frac{\mu(k)}{k^2} = \frac{6}{\pi^2}$. As the author writes, "These wonderful results are examples of Euler being Euler, manipulating symbols with a gusto that can take one's breath away. In so doing, he not only anticipated the Möbius function, but generated formulas more sophisticated than anything its namesake would discover eight decades later. Euler was, once again, far ahead of his time."

The article is written with its own gusto, guiding the reader with eloquence through the peculiar history of this foundational number-theoretic function. In a most entertaining way, "this tale reminds us—if we need reminding—that the history of mathematics can provide a host of unexpected rewards."

Response from the author

It is a thrill to receive the Allendoerfer Award for my article on the origins of the Möbius function. Many thanks to the MAA and to those committee members who directed this honor my way.

Let me share a little story. When Penny and I retired from Muhlenberg College in 2014, we moved to Bryn Mawr on Philadelphia's Main Line. This put us near Bryn Mawr College, and we were pleased when their mathematics department gave us an affiliation that let us enjoy a new academic home a few blocks from our real one.

As a historian, I especially appreciated the College's excellent Science Library. There, one might find an old book that once belonged to Charlotte Angas Scott, Bryn Mawr's first math professor, or a volume bearing the signature of Emmy Noether, the illustrious mathematician who was welcomed by Bryn Mawr after fleeing Nazi Germany in 1933.

One day my browsing led me to the collected works of August Ferdinand Möbius. I figured I'd thumb through it to find the famous Möbius function from the theory of numbers. But nothing inside smacked of number theory. It took some time before I spotted a version of the function buried within an 1832 paper on analysis. This suggested that there was more to the topic than meets the eye, a thought reinforced when I found that Leonhard Euler had stumbled upon the same function in his famous *Introductio in analysin infinitorum* from 1748. My attempts to unravel the history of this idea became an article for Mathematics Magazine. And here we are.

The take-away from my little tale: grazing through a great library can have unexpected rewards.

William Dunham is an historian of mathematics who has written/edited six books on the subject, including *Euler: The Master of Us All* (MAA, 1999) and *The Calculus Gallery* (Princeton, 2005). Since retiring from Muhlenberg College in 2014, he has held visiting positions at Princeton, Penn, Cornell, Harvard, and at Bryn Mawr College, where he is currently a research associate in Mathematics.

Jordan Bell and Victor Blåsjö

"Pietro Mengoli's 1650 Proof that the Harmonic Series Diverges." Mathematics Magazine Volume 91, Number 5, December 2018, pages 341–347.

Everyone who's taken enough calculus has seen at least one proof that the harmonic series diverges. Fewer know that Pietro Mengoli gave the first published proof of this divergence, in 1650. The result had been known for centuries, but Mengoli's argument had a distinctive flavor that merits a new look—particularly because modern accounts have not always faithfully relayed his methods. The authors take the reader on a journey not only through Mengoli's arguments, but through his actual words, by providing a complete English translation of the proof.

Mengoli's proof proceeds by grouping terms into blocks of three and using the inequality

$$\frac{1}{n-1}+\frac{1}{n}+\frac{1}{n+1}>\frac{3}{n}.$$

At this point it is tempting from our modern perspective to argue that we obtain a lower bound for the sum of the harmonic series in terms of the sum itself, and no finite quantity can satisfy such a bound. Indeed, this is what several modern accounts claim Mengoli's proof does.

The authors emphasize that Mengoli's proof did *not* do this—he was as wary of infinity as the ancient Greeks, from whom he drew direct inspiration. He took care to phrase his proof in finite terms: he applied the above inequality to blocks of 3 terms, then 9 terms, then 27 terms, and argued that by adding enough of these blocks, he could find a partial sum of the harmonic series that must exceed any given number. This is very much in keeping with Archimedes' approach to determining the area of a segment of a parabola, in which he evaluates a Riemann sum by ruling out every possible value for the area except for one.

This engaging article will draw in anyone who has thought about infinite series. One of its most appealing aspects is how it puts the reader in contact with Mengoli's original manuscript—the reader not only sees Mengoli's mathematics, but hears his voice. For instance, he begins with a meditation on Archimedes' determination of areas relating to parabolas. After summarizing Archimedes' argument, Mengoli interjects: "That wonderful theorem!" The reader comes away marveling at how mathematics can be a conversation across centuries.

Response from Bell and Blåsjö

We are delighted with this honor. It is very encouraging to see others share our excitement about the history of mathematics and our conviction that reflective engagement with the past has a natural place in current mathematical thought. As mathematicians with esoteric interests in history and philosophy, we are very fortunate to have such readers. Some forces in academia would rather push historians to a humanities department and hire another algebraic geometer in their place, but, thanks to the MAA, a more inclusive point of view is alive and well in the mathematical community. In this way, the thriving MAA community makes work such as ours possible through its promotion of a diversity of scholarly approaches to mathematics. We are immensely grateful for this invaluable support.

Viktor Blåsjö is an assistant professor at the Mathematical Institute of Utrecht University. He is a historian of mathematics with a special interest in the interplay between technical mathematical content and foundational issues in the early modern period. You can follow him on Twitter @viktorblasjo and listen to his Opinionated History of Mathematics podcast.

Jordan Bell is a mathematician and data scientist working in Toronto, Canada. His earlier work as a scholar in the history of mathematics includes a translation of Euler's paper finding the sum of reciprocals of the squares, and an exhaustive review paper on Euler's work on the pentagonal number theorem. He received his M.Sc. in mathematics from the University of Toronto. For the history of mathematics, Jordan's next project is a paper showing that the focus of Book I of Euclid's Elements is application of areas (I.44)—not the Pythagorean theorem (I.47)—and to present different medieval Latin proofs that fix a gap in Euclid's proof that has been seldom commented on in modern writings.