Vol. 89, No. 2, April 2016

MATHEMATICS MAGAZINE

- Magical finite projective planes
- Squigonometry for hyperellipses and supereggs
- Geometric phase from amusement rides to falling cats
- Testing the universality of quadratic forms

EDITORIAL POLICY

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

Submissions of articles are required via the *Mathematics Magazine*'s Editorial Manager System. The name(s) of the author(s) should not appear in the file. Initial submissions in pdf or LaTeX form can be sent to the editor at www.editorialmanager.com/mathmag/.

The Editorial Manager System will cue the author for all required information concerning the paper. Questions concerning submission of papers can be addressed to the editor at mathmag@maa.org. Authors who use LaTeX are urged to use the Magazine article template. However, a LaTeX file that uses a generic article class with no custom formatting is acceptable. The template and the Guidelines for Authors can be downloaded from www.maa.org/pubs/mathmag.

MATHEMATICS MAGAZINE (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, NW, Washington, DC 20036 and Lancaster, PA, in the months of February, April, June, October, and December.

Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Address advertising correspondence to

MAA Advertising
1529 Eighteenth St. NW
Washington, DC 20036
Phone: (202) 319-8461
E-mail: advertising@maa.org

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

Change of address, missing issue inquiries, and other subscription correspondence can be sent to:

The MAA Customer Service Center
P.O. Box 91112
Washington, DC 20090-1112
(800) 331-1622
(301) 617-7800
maaservice@maa.org

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

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, NW, Washington, DC 20036-1385.

Printed in the United States of America.

COVER IMAGE

Triples © 2016 David A. Reimann (*Albion College*). Used by permission.

The Proof Without Words "Infinitely Many Almost-Isosceles Pythagorean Triples Exist" by Roger Nelsen was the inspiration for this piece. Illustrated are primitive Pythagorean triangles with side leg length less than the arbitrary constant 180 and drawn with the hypotenuses of all triangles parallel to the main diagonal. The three triangles, (3, 4, 5), (20, 21, 29), and (119, 120, 169), shown below the main diagonal are almost-isosceles (their leg lengths differ by 1); the remaining triangles are shown above the main diagonal.

Vol. 89, No. 2, April 2016

MATHEMATICS MAGAZINE

LETTER FROM THE EDITOR

Once again the MAGAZINE offers expository articles on a wide assortment of mathematical topics. You are probably familiar with magic squares, but what about magic finite projective planes? David Nash and Jonathan Needleman ask and answer the question: When are finite projective planes magic?

A hyperellipse is a two-dimensional closed curve that is oval-like that was used by Danish designer Hein. In "Squigonometry, hyperellipses, and supereggs," Rob Poodiack considers a version on trigonometry based on hyperellipses instead of circles and along with calculus determines the area of the hyperellipse and the volume of the superegg, the solid of revolution of the hyperellipse.

Jeffrey Lawson and Matthew Rave use symmetry to analyze geometric phase in dynamical systems. Their examples are as engaging as they are diverse, including an amusement park ride and why cats (almost) always land on their feet.

Kenneth Williams shows how linear algebra can be used to bound the calculations necessary to determine if a positive quadratic form satisfies the conditions for two extraordinary theorems in number theory: the 15-Theorem and the 290-Theorem. These theorems give conditions for a quadratic form to be universal, meaning that every positive integer can be represented by the quadratic form.

Intransitive dice were brought to a broader audience by Martin Gardner in his Mathematical Games column in *Scientific American*. Brian Conrey, James Gabbard, Katie Grant, Andrew Liu, and Kent Morrison ask "How rare are intransitive dice?" They make conjectures about the frequency of certain types of dice behavior; these conjectures are supported by data. They also prove asymptotic results for the frequency of "one-step" dice to be intransitive.

There are also two proofs without words by Roger Nelsen and Raymond Viglione. Brendan Sullivan has constructed a mathematically oriented crossword puzzle about mathematical methods. As always, this issue has the ever popular Problems and Reviews sections.

Michael A. Jones, Editor

ARTICLES

When Are Finite Projective Planes Magic?

DAVID A. NASH
Le Moyne College
Syracuse, NY 13214
nashd@lemoyne.edu

JONATHAN NEEDLEMAN
Le Moyne College
Syracuse, NY 13214
needlejs@lemoyne.edu

Magic squares have a long history, with surviving written examples dating back to at least 300 BC. In fact, according to Chinese legends, a 3×3 magic square that is today known as the "Lo Shu" square was observed as a pattern on the shell of a tortoise by Emperor Yu sometime between 2200 and 2100 BC! Since those ancient times, magic squares have been marveled at and studied in numerous cultures. The idea is to fill in numbers into a square so that the sum along each row, column, and diagonal are all equal to the same number—often called the magic constant. For instance, the square in Figure 1 is a representation of the Lo Shu square. Here the magic constant is 15.

4	9	2
3	5	7
8	1	6

Figure 1 The Lo Shu magic square.

There have been numerous generalizations of magic squares to other shapes. Ely introduces the idea of magic designs [**4**]. A design is just a set of "points" and a set of "lines," with each line being a subset of points. A magic design is then an injective function from the points to the natural numbers where the sum along any line is constant. While Ely focused primarily on designs based on triangles and hexagons, other designs have since been studied. A particulary nice family of designs comes from (combinatorial) configurations. These are designs where every line has the same number of points, and every point has the same number of lines through it. Magic stars are an example with two lines through every point [**9**] and more recently Raney studied magic configurations where three lines pass through every point, and each line contains three points [**8**].

Projective planes are particularly nice configurations because the number of points on a line is the same as the number of lines through a point. Unfortunately, we will show that finite projective planes are never magic for any subset of the integers. Thankfully, there is no reason to limit ourselves to integers. All that is really needed to discuss

Math. Mag. **89** (2016) 83–91. doi:10.4169/math.mag.89.2.83. © Mathematical Association of America
MSC: Primary 51E15.

"magicness" is the ability to add the entries along a line. Because of this we consider making projective planes magic over Abelian groups.

For the special case of the Fano plane (the finite projective plane of order 2), Miesner and the first author [**6**] show that no magic labelings exist with labels in $\mathbb{Z}/n\mathbb{Z}$ for any n. In this paper we further generalize and study the "magicness" of all finite projective planes. Specifically, for every finite projective plane we will find a group for which it is magic and for a certain class of projective planes we will classify all groups for which they can be made magic.

For the interested reader, the authors further generalize the notion of magicness over Abelian groups to higher-dimensional finite projective spaces in [**7**].

Projective planes

We summarize some standard results about finite projective planes. For a general resource on finite projective planes the authors suggest [**1**] as an introduction and [**3**] for more advanced readers.

A projective plane $\Pi = (\mathcal{P}, \mathcal{L})$ is made up of a set of points $\mathcal{P}$ and a set of lines $\mathcal{L}$, where each line $L \in \mathcal{L}$ is a subset of $\mathcal{P}$. For any point $x \in \mathcal{P}$ we let $\mathcal{L}^x$ denote the set of all lines through x. To be a projective plane the following three axioms must hold.

1. There is a unique line between every pair of points—if $x, x' \in \mathcal{P}$ with $x \neq x'$, then there is a unique $L \in \mathcal{L}$ with $x, x' \in L$.
2. Each pair of lines intersects at a unique point—if $L, L' \in \mathcal{L}$ with $L \neq L'$, then there is a unique $x \in \mathcal{P}$ with $x \in L \cap L'$.
3. The plane contains a quadrilateral—there exist $x_1, x_2, x_3, x_4 \in \mathcal{P}$, so that there is no $L \in \mathcal{L}$ that contains any three of the points.

A finite projective plane $\Pi = (\mathcal{P}, \mathcal{L})$ is just a plane where the number of points $|\mathcal{P}|$ is finite. For finite projective planes the axioms imply some basic facts.

FACT. *For a finite projective plane $\Pi = (\mathcal{P}, \mathcal{L})$ there exists a number $n \in \mathbb{N}$, called the* order *of Π, and the following hold:*

1. *Every line contains $n + 1$ points, so that $|L| = n + 1$ for all $L \in \mathcal{L}$.*
2. *Every point is on $n + 1$ lines, so that $|\mathcal{L}^x| = n + 1$ for every $x \in \mathcal{P}$.*
3. *There are an equal number of points and lines, so that $|\mathcal{P}| = |\mathcal{L}| = n^2 + n + 1$.*

It is an open question to classify for which orders n there exists a projective plane, however, it is known that for any prime p and any $k \in \mathbb{N}$, there exists a projective plane of order $n = p^k$. In fact, these are the only orders for which projective planes are known to exist. However, finite projective planes have been completely classified for small orders and for sufficiently small order (≤ 8) they are all constructible in a uniform way. We will make use of that construction to deal with small order cases and thus we give it here.

Let $\mathbb{F}_q$ be a finite field of order $q = p^k$ for some prime p. We construct a projective plane $\Pi_q = (\mathcal{P}_q, \mathcal{L}_q)$ in the following way. View $\mathbb{F}_q^3$ as a vector space over $\mathbb{F}_q$, then let $\mathcal{P}_q = \{\text{1-dim subspaces of } \mathbb{F}_q^3\}$ and $\mathcal{L}_q = \{\text{2-dim subspaces of } \mathbb{F}_q^3\}$. One can verify that this construction yields a finite projective plane of order q. Since a point in Π_q is a line through the origin of $\mathbb{F}_q^3$ we can describe the points as follows. Given a nonzero $\langle x_1, x_2, x_3\rangle \in \mathbb{F}_q^3$, the set $[x_1, x_2, x_3] = \{\langle cx_1, cx_2, cx_3\rangle \mid c \in \mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}\}$ describes all points in $\mathcal{P}_q$. The lines in $\mathcal{L}_q$ are the planes through the origin in $\mathbb{F}_q^3$. Any vector v determines a plane through the origin by considering all vectors orthogonal to v. So the lines $\mathcal{L}_q$ are described as $[[x_1, x_2, x_3]] = \{u \in \mathbb{F}_q^3 \mid u \cdot \langle x_1, x_2, x_3\rangle = 0\}$, where $\langle x_1, x_2, x_3\rangle \in \mathbb{F}_q^3$ is a nonzero vector. Notice, for $c \in \mathbb{F}_q^*$, $[[cx_1, cx_2, cx_3]] = [[x_1, x_2, x_3]]$.

The smallest possible case when $q = 2$ is a plane with 7 points and 7 lines that is usually called the Fano plane, see Figure 2. There are 3 points on each line and 3 lines through each point. The Fano plane is actually the unique finite projective plane of order 2.

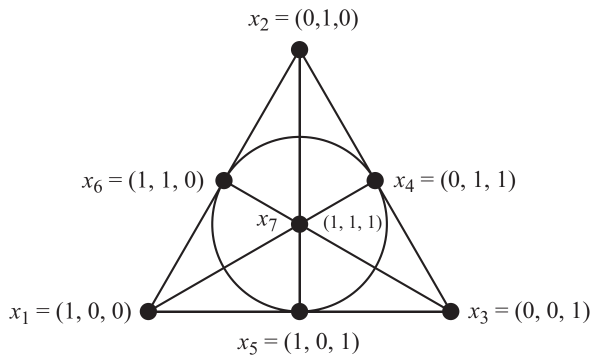


Figure 2 The Fano plane, Π_2.

Nonmagicness

An $n \times n$ square is magic if it is labeled with the numbers $\{1, \ldots, n^2\}$, so that each row, column, and diagonal sum to the same value. We can ask a similar question for a finite projective plane $\Pi = (\mathcal{P}, \mathcal{L})$ of order n. We would like to assign the values $\{1, \ldots, n^2 + n + 1\}$ to the points $\mathcal{P}$ so that the sum along any line is the same. Unfortunately, this is impossible, not only for the numbers $\{1, \ldots, n^2 + n + 1\}$, but for any set of $n^2 + n + 1$ distinct real numbers.

To prove this, we need the *incidence matrix* A of the projective plane Π. Let $x_1, \ldots, x_{n^2+n+1}$ be an enumeration of the points $\mathcal{P}$ and $L_1, \ldots, L_{n^2+n+1}$ be an enumeration of the lines $\mathcal{L}$. The rows of A will be indexed by the lines of Π and the columns by the points. Given $L_i \in \mathcal{L}$ and $x_j \in \mathcal{P}$, the entry $A_{i,j}$ is 1 if the point x_j is on the line L_i, and 0 otherwise. For example, working from Figure 2, if we take $L_1 = \overline{x_2x_3}$, $L_2 = \overline{x_1x_3}$, $L_3 = \overline{x_1x_2}$, $L_4 = \overline{x_1x_4}$, $L_5 = \overline{x_2x_5}$, $L_6 = \overline{x_3x_6}$, and $L_7 = \overline{x_4x_5}$, then the incidence matrix for Fano plane is as follows:

$$
\begin{array}{c} \\ L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \end{array}
\begin{array}{c} \begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \end{array} \\
\left[\begin{array}{ccccccc}
0 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 0
\end{array}\right] \end{array}.
$$

An important observation to make here is that the incidence matrix for any finite projective plane will be invertible over $\mathbb{R}$. A beautiful way to prove this fact is to consider the matrix AA^T. Observe the i, j-entry of the matrix AA^T is exactly the number of points on the intersection of the lines L_i and L_j. Since there are $n + 1$ points on every line, we can see that the diagonal entries of AA^T are all equal to $n+1$. In addition, distinct lines always intersect at a single point and hence all of the other entries in AA^T are equal to 1. One can then use row reduction to show that $\det(AA^T) = (n+1)^2 n^{n^2+n}$ (see, e.g., [**3**]). Thus AA^T, and more importantly A itself is invertible.

The incidence matrix also gives us a natural way of translating from a labeling of the points in $\mathcal{P}$ to the sums of those labels along each line in $\mathcal{L}$. To assist in this translation we view labelings as functions on the set of points $\mathcal{P}$. Let $f : \mathcal{P} \to \mathbb{R}$ denote a function so that the sum along each line in $\mathcal{L}$ is the same magic constant $c \in \mathbb{R}$. If such an f exists, which is also injective, then we say that Π is *magic over* $\mathbb{R}$. The magic constant condition can then be represented by the matrix equation $A\mathbf{f} = \mathbf{c}$, where A is the incidence matrix for Π, $\mathbf{f}$ is the column vector $[f(x)]_{x\in\mathcal{P}}$, and $\mathbf{c}$ is the vector with all entries equal to c. Using this data, we demonstrate that the plane Π is not magic over $\mathbb{R}$.

Proposition 1. *No finite projective plane $\Pi = (\mathcal{P}, \mathcal{L})$ is magic over $\mathbb{R}$.*

Proof. Let $f : \mathcal{P} \to \mathbb{R}$ be any real valued function on the points $\mathcal{P}$. $\mathbf{f}$ is a column vector with the rows indexed by $\mathcal{P}$. If A is the incidence matrix for Π, then $A\mathbf{f}$ will be a column vector with the rows indexed by $\mathcal{L}$. Hence we may think of $A\mathbf{f}$ as a real-valued function on $\mathcal{L}$. The value of the row indexed by $L \in \mathcal{L}$ is exactly $\sum_{x\in L} f(x)$, the sum of f along L, by construction of A.

We are interested in when $A\mathbf{f}$ is a constant function. However, as mentioned above, the incidence matrix is invertible. This means that the equation $A\mathbf{f} = c$ has a unique solution which, therefore, must be the constant function $f(x) = \frac{c}{n+1}$ for all $x \in \mathcal{P}$. Since $\mathbf{f}$ is not injective, it follows that Π is not magic over $\mathbb{R}$. ■

In light of the result above, we seek to find conditions under which a finite projective plane $\Pi = (\mathcal{P}, \mathcal{L})$ can be considered magic. We must be able to add the values assigned to points and the addition must be commutative since the points are not in any particular order. So we let G be an Abelian group and consider a G-valued function $v : \mathcal{P} \to G$ on the points of a projective plane. For any subset $S \subset \mathcal{P}$ we then define $v(S) = \sum_{x\in S} v(x)$. The function v is called *line invariant* if $v(L) = v(L')$ for all $L, L' \in \mathcal{L}$. When this holds we call $v(L)$ the *magic constant*.

The set of line invariant functions has a really beautiful structure, but unfortunately, it includes the constant functions which are trivially line invariant because every line has the same number of points. Since constant functions do not get at the nature of magicness, we will only refer to v as a *pseudomagic* function when it is both line invariant and nonconstant. If, furthermore, v is actually injective, then we will call v *magic*.

We say a finite projective plane $\Pi = (\mathcal{P}, \mathcal{L})$ is *pseudomagic over an Abelian group* G (resp. *magic over* G) and G *admits a pseudomagic* (resp. *magic*) *function* v if and only if there exists a pseudomagic (resp. magic) function $v : \mathcal{P} \to G$. As we will show in Theorem 2, Π will not admit a pseudomagic function over a group G unless that group contains elements of finite order. Groups that do not contain any elements of finite order are called *torsion-free* groups.

Theorem 2. *Let G be an Abelian torsion-free group and let $\Pi = (\mathcal{P}, \mathcal{L})$ be a finite projective plane. Then Π is not pseudomagic over G.*

Proof. Let v be a line invariant G-valued function on $\mathcal{P}$. Let H be the subgroup of G generated by the values v take on $\mathcal{P}$, so that $H = \langle v(x) \mid x \in \mathcal{P}\rangle$. H is a finitely generated torsion-free Abelian group and thus $H \cong \mathbb{Z}^k$ for some $k \in \mathbb{N}$ (see, e.g., [**2**]). We may therefore view v as a $\mathbb{Z}^k$-valued function, and we may write

$$v(x) = (v_1(x), v_2(x), \dots, v_k(x)) \in \mathbb{Z}^k,$$

so that each component function v_i maps $\mathcal{P} \to \mathbb{Z}$. Since $v(L)$ is independent of $L \in \mathcal{L}$, it also holds that $v_i(L)$ is independent of L for each i. By Proposition 1 each v_i is a constant function and hence v is constant on $\mathcal{P}$ as well. ■

Magicness

In order to find a group over which a projective plane is magic we must look at torsion groups. Cyclic groups are a natural place to start. We begin by classifying cyclic groups that admit a pseudomagic function for a given projective plane. Throughout the section let $\Pi = (\mathcal{P}, \mathcal{L})$ be a projective plane of order n and let $m \in \mathbb{N}$.

One way to attempt to find a pseudomagic function from $\mathcal{P}$ to $\mathbb{Z}/m\mathbb{Z}$ is to create a function that is constant on a chosen line L and zero on all points off of L. As it turns out, if the constant is chosen carefully, then the function will be line invariant. More precisely, given any line $L \in \mathcal{L}$, we define the function $v_L : \mathcal{P} \to \mathbb{Z}/m\mathbb{Z}$ as follows:

$$v_L(x) = \begin{cases} \frac{m}{(n,m)} & \text{if } x \in L \\ 0 & \text{if } x \notin L \end{cases}, \tag{1}$$

where (n, m) is the greatest common divisor of n and m. Notice that when m divides n, v_L is the characteristic function of L. As an example, consider the Fano plane (of order $n = 2$) and the group $G = \mathbb{Z}/6\mathbb{Z}$ (so that $m = 6$). We may choose the line L_1, then the function v_{L_1} would correspond to labeling the points on that line by 3, see Figure 3.

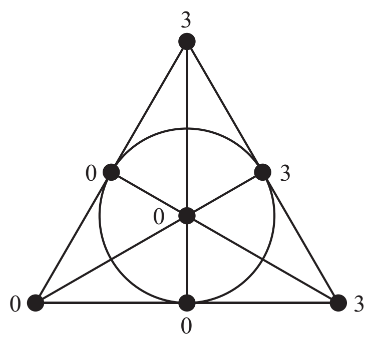


Figure 3 The pseudomagic function v_{L_1} from the points in the Fano plane to $\mathbb{Z}/6\mathbb{Z}$.

The function v_{L_1} is line invariant because the sum along any line is $3 \in \mathbb{Z}/6\mathbb{Z}$. We now show more generally that the functions v_L for each $L \in \mathcal{L}$ are always line invariant. This fact relies directly on the incidence structure of finite projective planes.

Lemma 3. *$v_L : \mathcal{P} \to \mathbb{Z}/m\mathbb{Z}$ is line invariant.*

Proof. Observe that we have

$$v_L(L) = \sum_{x \in L} \frac{m}{(n,m)} \equiv \frac{(n+1)m}{(n,m)} \equiv \frac{nm}{(n,m)} + \frac{m}{(n,m)} \mod m. \tag{2}$$

However, because m divides $\frac{nm}{(n,m)}$, we have $v_L(L) = \frac{m}{(n,m)} \mod m$. Let $L' \in \mathcal{L}$ be any other line. We know that L' intersects L in exactly one point x. Thus, $v_L(x) = \frac{m}{(n,m)}$ and $v_L(x') = 0$ for all other $x' \in L'$. It follows that $v_L(L') = \frac{m}{(n,m)}$ as well. ■

Corollary 4. *If* $(n, m) > 1$*, then* Π *is pseudomagic over* $\mathbb{Z}/m\mathbb{Z}$*.*

Proof. Let v_L be defined as in Equation 1 for any line L. Then, by definition $v_L = 0$ if and only if $(n, m) = 1$. Thus, when $(n, m) \neq 1$, the line invariant function v_L is nonzero and is therefore a pseudomagic function on $\mathcal{P}$ for each line $L \in \mathcal{L}$. ■

It turns out these are the only cyclic groups for which Π can be made pseudomagic.

Proposition 5. Π *is pseudomagic over* $\mathbb{Z}/m\mathbb{Z}$ *if and only if* $(n, m) \neq 1$*. Furthermore,* Π *is never magic over* $\mathbb{Z}/m\mathbb{Z}$*.*

Proof. The "if" is proven in Corollary 4. For the other direction assume there is a line invariant function $v : \mathcal{P} \to \mathbb{Z}/m\mathbb{Z}$ with magic constant $g \in \mathbb{Z}/m\mathbb{Z}$. Let $a, b \in \mathcal{P}$ and let $L = \overline{ab}$ be the line containing a and b and let L^c be the set of points not on L. Each point of L^c is on exactly one line in the set of lines $\mathcal{L}^a \setminus \{L\}$ (recall $\mathcal{L}^a$ is all lines through a). Therefore,

$$v(L^c) = \sum_{L' \in \mathcal{L}^a \setminus \{L\}} (v(L') - v(a)) = ng - nv(a),$$

since there are n lines in $\mathcal{L}^a \setminus \{L\}$ and a is not in L^c. Similarly, $v(L^c) = ng - nv(b)$ and hence $nv(a) = nv(b)$ in $\mathbb{Z}/m\mathbb{Z}$. Since a, b were arbitrary, we may conclude that $nv(x)$ is independent of $x \in \mathcal{P}$. Therefore, $n(v(x) - v(y)) = 0$ for all $x, y \in \mathcal{P}$. Hence, for all $x, y \in \mathcal{P}$, $v(x) - v(y)$ is in the kernel of the homomorphism $\phi_n : t \mapsto nt$ in $\mathbb{Z}/m\mathbb{Z}$, and so each $v(x)$ is in the same coset of the kernel. The homomorphism has $|\mathrm{Ker}(\phi_n)| = (n, m)$, and so a line invariant function $v : \mathcal{P} \mapsto \mathbb{Z}/m\mathbb{Z}$ can take on up to (n, m) different values. Thus, when $(n, m) = 1$, v is a constant function and not a pseudomagic function. Furthermore, $(n, m) < n^2 + n + 1 = |\mathcal{P}|$, so v can never be magic. ■

From Corollary 4, it is not hard to find a group G which admits a magic function on Π. Take one copy of $\mathbb{Z}/n\mathbb{Z}$ for every line in $\mathcal{L}$. That is let $G = (\mathbb{Z}/n\mathbb{Z})^k$, where $k = |\mathcal{L}| = n^2 + n + 1$. Next, choose an enumeration of $\mathcal{L} = \{L_1, L_2, \ldots, L_k\}$ and define $v : \mathcal{P} \to (\mathbb{Z}/n\mathbb{Z})^k$ as $v(x) = (v_{L_1}(x), v_{L_2}(x), \ldots, v_{L_k}(x))$. By Corollary 4 this is a pseudomagic function. Recall that to be magic this function must also be injective. The injectivity of v follows from the fact that for any two points $x_1, x_2 \in L$ there exists a line that contains one of them but not the other.

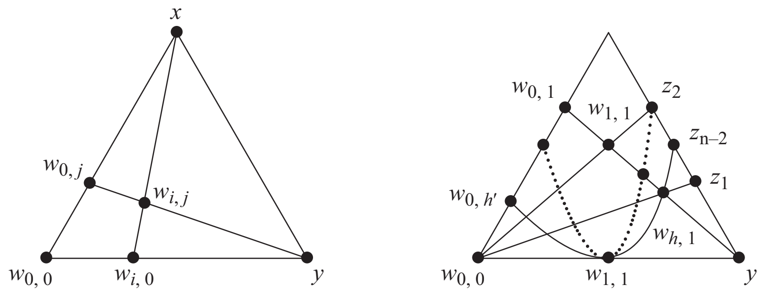
This construction seems inefficient as $|G|$ is much larger than $|\mathcal{P}|$ and the exponent k depends on Π. Next, we find the smallest r so that Π is magic over $(\mathbb{Z}/n\mathbb{Z})^r$. Since $|\mathcal{P}| = n^2 + n + 1$ we know $r \geq 3$, and in fact, we show $r = 3$ works. Our general proof relies on having $n \geq 5$ and hence we treat the cases $n = 2$, 3, and 4 separately.

Theorem 6. *If* $\Pi = (\mathcal{P}, \mathcal{L})$ *is a projective plane of order* $n \geq 5$*, then* Π *is magic for the group* $G = (\mathbb{Z}/n\mathbb{Z})^3$*.*

Proof. The plan is to create three separate pseudomagic functions v_1, v_2, and v_3 from $\mathcal{P}$ to $\mathbb{Z}/n\mathbb{Z}$ which together define a magic function $(v_1, v_2, v_3) : \mathcal{P} \to G$.

To begin, we label some points and lines for reference. Let x be any point in $\mathcal{P}$ and let $L_0, L_1, \ldots, L_n$ be an enumeration of the $n + 1$ lines through x. Next, let y be another point on L_n and let $L'_0, \ldots, L'_{n-1}$ denote the other n lines through y. Then, for each $1 \leq i, j \leq n - 1$, we let $w_{i,j} = L_i \cap L'_j$ as in Figure 4a. Finally, let $z_1, z_2, \ldots, z_{n-1}$ denote the points in $L_n \setminus \{x, y\}$. For now the choice of the z_k is arbitrary, but later we will be more specific in our labeling.

From this point forward a couple of minor details in the proof depend on the parity of n. The main technical difference stems from the fact that the sum of all of the elements in $\mathbb{Z}/n\mathbb{Z}$ is 0 when n is odd, but is $\frac{n}{2}$ when n is even. As it turns out, our



(a) (b)

Figure 4 Labeling $\mathcal{P}$.

construction is unaffected by this difference, but for simplicity we choose to deal with n odd first, and then explain why it works for n even as well.

For n odd define the first two pseudomagic functions as follows:

$$v_1 = \sum_{k=1}^{n}(k-1)v_{L_k}, \quad v_2 = \sum_{k=1}^{n}(k-1)v_{L'_k}. \tag{3}$$

Since x is on each L_k we have $v_1(x) = \sum_{k=1}^{n}(k-1) = 0$ in $\mathbb{Z}/n\mathbb{Z}$. Similarly, $v_2(y) = 0$. The function $v' := (v_1, v_2) : \mathcal{P} \to (\mathbb{Z}/n\mathbb{Z})^2$ then already has unique values on most of the points. The only equalities are the following:

$$\begin{array}{ll} v'(w_{0,0}) = v'(w_{0,1}) = v'(w_{1,0}) = v'(w_{1,1}) = (0,0) & \\ v'(w_{0,i}) = v'(w_{1,i}) = (0, i-1) & 2 \le i \le n-1 \\ v'(w_{i,0}) = v'(w_{i,1}) = (i-1, 0) & 2 \le i \le n-1 \\ v'(z_i) = v'(z_j) = (n-1, n-1) & 1 \le i, j \le n-1 \end{array} \tag{4}$$

since $v'(w_{i,j}) = (i-1, j-1)$ for $1 \le i, j \le n-1$.

We now create a third function v_3 to distinguish points which have equal values on v'. This is a delicate construction which requires a careful ordering of the points z_k with respect to other points on the plane. Let $z_2 = L_n \cap \overline{w_{0,0}w_{1,1}}$ and let J' be the line $\overline{w_{1,0}z_2}$. Next, choose an h so that $w_{h,1} \in L'_1 \backslash \{w_{0,1}, y, w_{1,1}, J' \cap L'_1\}$. This is possible since $n \ge 5$ and so there must be at least 6 points on a line. Now let $z_1 = \overline{w_{0,0}w_{h,1}} \cap L_n$, and $z_{n-2} = \overline{w_{1,0}w_{h,1}} \cap L_n$. Once again, since $n \ge 5$, $z_{n-2} \neq z_2$. Label the remaining points on L_n as $\{z_3, \dots z_{n-3}, z_{n-1}\}$, and let $L''_i = \overline{w_{0,0}z_i}$ for $1 \le i \le n-1$. Define the line $J = \overline{w_{1,0}z_{n-2}}$. For future reference we will let $w_{0,h'} = L_0 \cap J$. Figure 4b depicts the specific labeling that we have described with J' dashed and J solid.

We now define $v_3 : \mathcal{P} \to \mathbb{Z}/n\mathbb{Z}$ by

$$v_3 = \sum_{k=1}^{n-1}(k-1)v_{L''_k} + (n-1)v_{L_0} + 2v_J. \tag{5}$$

All that is left is to check that v_3 differentiates the points that had equal values under v'. First, check the z_k. Observe that $v_3(z_k) = k-1$ if $1 \le k \le n-1$ with $k \neq n-2$, and $v_3(z_{n-2}) = n-1$. The difference for z_{n-2} is the inclusion of $2v_J$ in the definition of v_3.

Next, check the pairs $w_{i,0}$ and $w_{i,1}$. For $2 \le i \le n-1$, $v_3(w_{i,0}) = 0$. For each of those i except h there is a unique j with $2 \le j \le n-1$ so that $w_{i,1} \in L''_j$. In these cases $v_3(w_{i,1}) = j - 1 \neq 0$. For h, $w_{h,1} = L''_1 \cap J$, so $v_3(w_{h,1}) = 2$.

Now, check the pairs $w_{0,i}$ and $w_{1,i}$. For each $1 \le i \le n-1$ there is a unique j with $1 \le j \le n-1$ so that $w_{1,i} \in L''_j$, we therefore have $v_3(w_{1,i}) = j - 1 \neq n-1$. Recall,

h' is defined so that $w_{0,h'} = J \cap L_0$. For $1 \le i \le n-1$ with $i \ne h'$ we have $v_3(w_{0,i}) = n-1$, and $v_3(w_{0,h'}) = 1$. We need to show $v_3(w_{1,h'}) \ne 1$. We know $v_3(w_{1,1}) = 1$ since $w_{1,1} \in L''_2$. If $h' = 1$, then $J = \overline{w_{0,1}w_{h,1}}$, but $w_{1,0} \in J$ and $w_{1,0} \notin \overline{w_{0,1}w_{h,1}}$, so $h' \ne 1$.

Finally, check $w_{0,0}$, $w_{0,1}$, $w_{1,0}$, and $w_{1,1}$. Observe that $v_3(w_{0,1}) = n-1$, $v_3(w_{1,0}) = 2$, $v_3(w_{1,1}) = 1$, and $v(w_{0,0}) = 0$.

Thus $v = (v_1, v_2, v_3) : \mathcal{P} \to (\mathbb{Z}/n\mathbb{Z})^3$ is injective and hence magic when n is odd. In the case when n is even define v_1, v_2, v_3 in the same fashion, but now $v_1(x) = v_2(y) = v_3(w_{0,0}) = n/2$. Since $n \ge 5$ we have $2 < \frac{n}{2} < n-1$ and thus these changes do not impact the proof. ■

We now treat the cases $n = 2$, 3, and 4 separately.

Proposition 7. *The projective planes of order $n = 2$, 3, and 4 are magic over $G = (\mathbb{Z}/n\mathbb{Z})^3$.*

Proof. For each $2 \le n \le 4$ there is a unique projective plane of order n [**5**] and hence we may use the construction of the plane Π_n described earlier. Recall that each point $a \in \mathcal{P}_n$ can be represented in the form $a = [x, y, z] = \{\langle cx_1, cx_2, cx_3\rangle \mid c \in \mathbb{F}_q^*\}$ for some nonzero vector $\langle x, y, z\rangle \in \mathbb{F}_n^3$. In the $n = 2$ case we have the Fano plane (see Figure 2) and the function

$$v(a) = (x, y, z)$$

is magic. In the $n = 3$ case, define the following functions:

$$\begin{aligned} v_1(a) &= x^2 + z^2 + xy + 2yz + 2xz \\ v_2(a) &= y^2 + x^2 + yz + 2zx + 2yx \\ v_3(a) &= z^2 + y^2 + zx + 2xy + 2zy \end{aligned}$$

and then $v = (v_1, v_2, v_3)$ is magic.

These functions are well defined on the sets $[x, y, z]$, since the components are homogeneous polynomials of degree $n-1$ and for any $c \in \mathbb{F}_n$ we have $c^{n-1} = 1$. One can directly check these that these functions are magic.

For $n = 4$ we need a significant modification of our construction from Theorem 6. We are working with $\mathbb{F}_4 = \{0, 1, \alpha, \alpha+1\}$, where $\alpha^2 = \alpha + 1$ and we set $a_0 = 0$, $a_1 = 1$, $a_2 = \alpha$, and $a_3 = \alpha + 1$. We then let $w_{i,0} = [1, 0, a_i]$, $w_{0,i} = [1, a_i, 0]$, and $z_i = [0, 1, a_i]$ for $0 \le i \le 3$, and we let $y = [0, 0, 1]$ and $x = [0, 1, 0]$. Next we define lines $L_i = \overline{xw_{i,0}}$, $L'_i = \overline{yw_{0,i}}$, $L''_i = \overline{w_{0,0}z_i}$, and the rest of the points $w_{i,j}$ as in Theorem 6. Using these lines we again define v_1 and v_2 as in (3). This leaves us with the same equalities described in (4). We define the line $J = \overline{y_1 w_{2,1}}$ and then set

$$v_3 = 3v_{L_0} + v_{L''_2} + 2v_{L''_3} + v_J.$$

One can then directly check that $v = (v_1, v_2, v_3)$ is magic. ■

To this point we have a relatively small group for which a projective plane is magic, however in the case when n is a prime we can say more. It turns out that in some sense $(\mathbb{Z}/n\mathbb{Z})^3$ is the only group for which Π is magic.

Theorem 8. *If n, the order of Π is prime and Π is magic over some Abelian group G, then $(\mathbb{Z}/n\mathbb{Z})^3$ is a subgroup of G.*

Proof. Let $v : \mathcal{P} \to G$ be a magic function. We may assume G is finitely generated since we may work with the subgroup generated by Im v. Since G is finitely generated there exist $n_1, \ldots, n_k \in \mathbb{N}$ such that $G = \bigoplus_{i=1}^k \mathbb{Z}/n_i\mathbb{Z}$. Thus $v = \bigoplus_{i=1}^k v_i$, where $v_i : \mathcal{P} \to \mathbb{Z}/n_i\mathbb{Z}$ is the natural projection of v into $\mathbb{Z}/n_i\mathbb{Z}$, and each v_i is line invariant. Without loss of generality we assume each v_i is nonconstant and hence pseudomagic.

Since n is prime, Proposition 5 implies that $n \mid n_i$, and furthermore, the proof shows $|\text{Im } v_i| \leq (n, n_i) = n$. This means $k \geq 3$ as otherwise $|\text{Im } v| \leq n^2 < |\mathcal{P}|$ which contradicts v being injective. ■

When n is not prime it might be possible for smaller groups to admit a magic function. If m divides n and $m > 1$, then $mn^2 > n^2 + n + 1 = |\mathcal{P}|$, so Π could potentially be magic over a group of order mn^2.

OPEN QUESTIONS

1. *When m divides n with $m > 1$, is Π magic over $(\mathbb{Z}/n\mathbb{Z})^2 \times \mathbb{Z}/m\mathbb{Z}$?*

As it turns out, for planes of the form Π_q when q is prime, every line invariant function to $\mathbb{Z}/m\mathbb{Z}$ is a linear combination of the v_L functions. However, this is not true when q is not prime (see [**7**]). In the nonprime case there are special line invariant functions which are not linear combinations of the v_L's. Our proof of Theorem 6 does not make use of these special functions. Therefore, it seems reasonable to expect that if our open question is answered in the affirmative, then the proof will require the special functions.

Another question that one could ask is whether or not all magic functions to larger groups G actually come from functions to $(\mathbb{Z}/n\mathbb{Z})^3$. A more technical way to describe this is given below.

2. *Suppose that $v : \Pi \to G$ is a magic function and Π is order n, under what conditions is there a surjection $f : G \to (\mathbb{Z}/n\mathbb{Z})^3$ such that $f \circ v : \Pi \to (\mathbb{Z}/n\mathbb{Z})^3$ is still magic?*

Acknowledgments The second author would like to thank Michael Raney for originally introducing him to magic configurations. Both authors would also like to thank the referees and the editor for their helpful suggestions.

REFERENCES

1. A. A. Albert, R. Sandler, *An Introduction to Finite Projective Planes*. Holt, Rinehart and Winston, New York, 1968.
2. M. A. Armstrong, *Groups and Symmetry*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1988.
3. P. Dembowski, *Finite Geometries*. Reprint of the 1968 original, Classics in Mathematics. Springer-Verlag, Berlin, 1997.
4. R. B. Ely III, Magic designs, *J. Recreat. Math.* **1** no. 1 (1968) 3–17.
5. C. R. MacInnes, Finite planes with less than eight points on a line, *Amer. Math. Monthly* **14** (1907) 171–174.
6. B. Miesner, D. Nash, Magic Fano planes Π*ME J.* **14** no. 1 (2014) 23–29.
7. D. Nash, J. Needleman, On magic finite projective space, http://arxiv.org/abs/1412.1545.
8. M. W. Raney, On geometric trilateral-free (n_3) configurations, *Ars Math. Contemp.* **6** no. 2 (2013) 253–259.
9. M. Trenkler, Magic stars, Π*ME J.* **11** no. 10 (2004) 549–554.

Summary. We study a generalization of magic squares, where the entries come from the natural numbers, to magic finite projective planes, where the entries come from Abelian groups. For each finite projective plane we demonstrate a small group over which the plane can be labeled magically. In the prime order case we classify all groups over which the projective plane can be made magic.

DAVID A. NASH (MR Author ID:924244) earned his Ph.D. from University of Oregon in 2010. Since then he has been working as an assistant professor of mathematics at Le Moyne College. He enjoys sharing his passion for mathematics and problem solving with his students. While his primary research areas center predominantly on representation theory and category theory, he has especially appreciated involving undergraduates in research more recently by tackling interesting problems with more recreational origins.

JONATHAN NEEDLEMAN (MR Author ID:763644) earned his Ph.D. from Cornell University in 2009. After spending a year as a visiting professor at Oberlin College he became an assistant professor of mathematics at Le Moyne College. He has an unhealthy obsession with puzzles, games, and recreational mathematics, and enjoys helping people find a love of math through play.

Squigonometry, Hyperellipses, and Supereggs

ROBERT D. POODIACK
Norwich University
Northfield, VT 05663
rpoodiac@norwich.edu

Piet Hein (1905–1996) was a man of many interests: a scientist, a poet, a philosopher. He made an indelible mark in mathematics, though, as a designer. Hein noticed that many of the objects of modern life encompassed circles or rectangles, and devoted time to determine which closed curve would be a happy medium, a curve that would incorporate the most pleasing aspects of both [**3**, Chapter 6].

Hein, a Dane, addressed this problem at the request of urban planners in Stockholm, Sweden in 1959. The planners were attempting to flow traffic through a rectangular plot of land, in the center of which was an oval-shaped fountain. Hein's solution was to lay out the street surrounding the fountain in the shape of a hyperellipse (or as Hein called it, a "superellipse"). The resulting central square, Sergels Torg, is considered to be an example of harmonious city design. (See Figure 1.)

Figure 1 Sergels Torg in Stockholm, Sweden. Photograph by Anders Bengtsson, distributed under a CC-BY 2.0 license.

The hyperellipse—a curve with equation $|x/a|^p + |y/b|^p = 1$, $p \geq 1$—became a model for Scandinavian architectural and furniture design throughout the 1960s. (Hein used $p = 2.5$ in his design of Sergels Torg [**3**, p. 66].) The solid of revolution for a hyperellipse with $p > 2$, dubbed a "superegg," became a *cause celèbre*. The superegg has zero curvature on its top and bottom, and can sit on a table without toppling, unlike an ellipsoid. At the website named for Hein [**5**], one can obtain small supereggs to chill drinks, to be manipulated for stress relief, and as sculptures to dress up a room.

Math. Mag. **89** (2016) 92–102. doi:10.4169/math.mag.89.2.92.
MSC: Primary 00, Secondary 14; 34

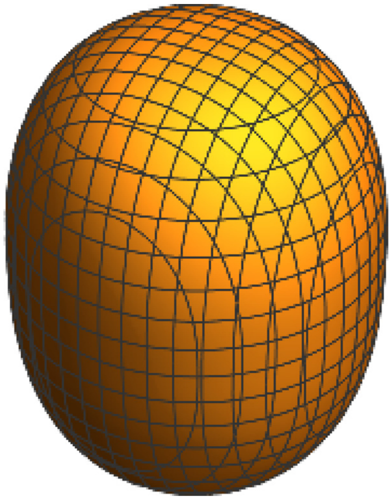


Figure 2 Superegg with $p = 2.5$.

What is the area of a hyperellipse? What is the volume of its associated superegg? We can determine these by extending the ideas of trigonometry—recently dubbed "squigonometry" [**11**]—and calculus to these curves and solids.

Review of squigonometric functions Many mathematicians have explored generalizing trigonometric functions ([**4, 6, 9**]) in various directions over the years. Squigonometry is so called because the trigonometric functions obtained represent the coordinates of points on a unit *squircle* with equation $|x|^p + |y|^p = 1$. The resulting curve is indeed a squarish circle for $p > 2$.

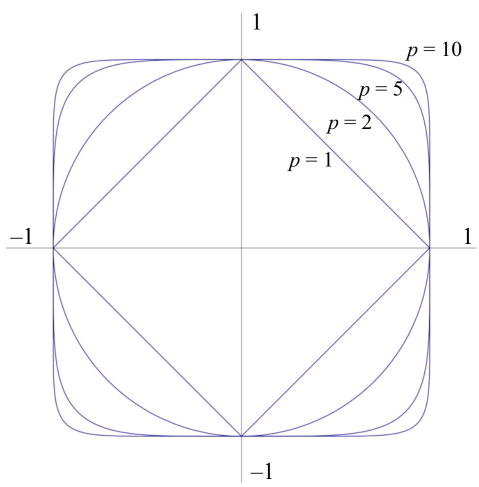


Figure 3 Unit squircles $|x|^p + |y|^p = 1$ for $p = 1, 2, 5,$ and 10.

The cosquine and squine functions, $\text{cq}_p(t)$ and $\text{sq}_p(t)$, are defined as the unique solutions to the coupled initial value problem

$$\begin{cases} x'(t) = -y(t)^{p-1} \\ y'(t) = x(t)^{p-1} \\ x(0) = 1 \\ y(0) = 0, \end{cases} \tag{1}$$

where x corresponds to the cosquine, and y to the squine.

Combination and integration of the system of equations (1) yields the relation

$$\mathrm{sq}_p^p(t) + \mathrm{cq}_p^p(t) = 1. \tag{2}$$

For $p = 4$, the value explored in [**11**], or $p = 2$, the customary and usual value, or any even p, we get the full unit squircle (or circle). However, as was pointed out in Exercise 11 of [**11**], we can achieve the full squircle for an arbitrary $p \geq 1$, as in Figure 3, by simply taking absolute values ($|x|^p + |y|^p = 1$) or by deftly restricting the interval of solution for the system of equations (1) and then extending the solutions by symmetry. We will see how to do this presently.

General versions of the inverse squine and cosquine defined in [**11**] can be derived from equations (1) and (2). Let $x = \mathrm{sq}_p(y)$. Then $dx/dy = \mathrm{cq}_p^{p-1}(y) = (1 - x^p)^{(p-1)/p}$. Separating and integrating yields

$$y = \mathrm{sq}_p^{-1}(x) = \int_0^x \frac{1}{(1-t^p)^{(p-1)/p}}\,dt. \tag{3}$$

In a similar manner, we have

$$\mathrm{cq}_p^{-1}(x) = \int_x^1 \frac{1}{(1-t^p)^{(p-1)/p}}\,dt. \tag{4}$$

The many roles of π The above inverse squigonometric functions lead organically to the definition of a more general version of π. We define π_p to be

$$\pi_p = 2\int_0^1 \frac{1}{(1-t^p)^{(p-1)/p}}\,dt = 2\,\mathrm{sq}_p^{-1}(1) = 2\,\mathrm{cq}_p^{-1}(0). \tag{5}$$

This means that $\mathrm{sq}_p(\pi_p/2) = 1$ and $\mathrm{cq}_p(\pi_p/2) = 0$. In addition, because there are only four lines of symmetry for the unit squircle [**11**], we have that $\mathrm{sq}_p(\pi_p/4) = \mathrm{cq}_p(\pi_p/4) = 1/\sqrt[p]{2}$, which corresponds to the upper right-hand "corner point" of the squircle.

It follows from equations (3), (4), and (5) that $\mathrm{sq}_p^{-1}(x) + \mathrm{cq}_p^{-1}(x) = \pi_p/2$ for $-1 \leq x \leq 1$ and that we get the expected ranges:

$$-\pi_p/2 \leq \mathrm{sq}_p^{-1}(x) \leq \pi_p/2,$$
$$0 \leq \mathrm{cq}_p^{-1}(x) \leq \pi_p.$$

We will soon show in an example that the area under the first-quadrant portion of the unit p-circle is given by $\pi_p/4$ square units. From this and numerical integration, we see in Figure 5 that the value of π_p increases from $\pi_1 = 2$ toward a limiting value of 4, four times the area of the unit square $[0, 1] \times [0, 1]$, as $p \to \infty$. (An 1897 bill introduced in the Indiana State Legislature resolved to set the value of π to 4. The Biblical value for π of 3 is achieved when $p \approx 1.79148$ [**2**]. The value of π_2 is, of course, our familiar π.) Note that we can think of $\pi_p/2$ as being the measure of a right angle. See [**1, 2, 9**] for additional generalizations of π and their interpretations.

p	π_p
1	2.000000000
2	3.141592654
3	3.533277501
4	3.708149355
5	3.800600556
6	3.855242593
7	3.890174723
8	3.913843115
9	3.930613253
10	3.942922898

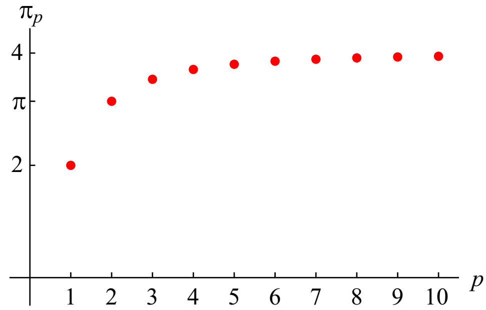


Figure 4 π_p values via numerical integration.

We can express π_p in terms of the beta or gamma functions. Recall that the beta function is defined as

$$B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1}\,du,$$

and its relationship to the gamma function is via the identity $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ and $\Gamma(n) = (n-1)!$ for n a positive integer.

If we make the change of variable $u = t^p$ in equation (5), we get

$$\pi_p = \frac{2}{p}\int_0^1 u^{1/p-1}(1-u)^{1/p-1}\,du = \frac{2}{p}B(\tfrac{1}{p}, \tfrac{1}{p}) = \frac{2}{p}\frac{[\Gamma(\frac{1}{p})]^2}{\Gamma(\frac{2}{p})}.$$

We can now determine exact values for some π_p's by use of well-known identities for the gamma function:

$$\Gamma(x+1) = x\Gamma(x) \quad \text{and} \quad \Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}. \tag{6}$$

For instance,

$$\pi_3 = \frac{2[\Gamma(\frac{1}{3})]^2}{3\Gamma(\frac{2}{3})} = \frac{[\Gamma(\frac{1}{3})]^3}{\sqrt{3}\,\pi} \quad \text{and} \quad \pi_4 = \frac{[\Gamma(\frac{1}{4})]^2}{2\Gamma(\frac{1}{2})} = \frac{[\Gamma(\frac{1}{4})]^2}{2\sqrt{\pi}}. \tag{7}$$

These formulas imply that $\Gamma(\frac{1}{3}) = \sqrt[3]{\pi_3\pi\sqrt{3}}$ and $\Gamma(\frac{1}{4}) = \sqrt{2\pi_4\sqrt{\pi}}$. A general formula for $\Gamma(\frac{1}{2^n})$ involving π_{2^k}, $1 \le k \le n$ is found in [**9**].

Matters are more difficult for other values of p. For example, we cannot solve for π_5 in terms of a single gamma function value, but we can use (6) to help derive an expression for finding the volume of the superegg with $p = 2.5$:

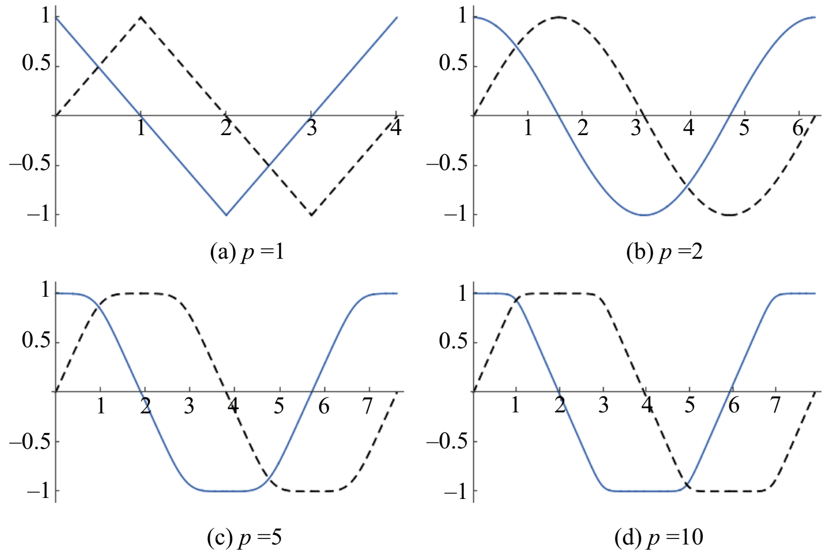


Figure 5 Graphs of $y = \mathrm{cq}_p(t)$ (solid) and $y = \mathrm{sq}_p(t)$ (dashed) for various values of p.

$$\begin{aligned}\pi_5 &= \frac{2[\Gamma(\frac{1}{5})]^2}{5\Gamma(\frac{2}{5})} = \frac{2}{5}\frac{\Gamma(\frac{1}{5})}{\Gamma(\frac{2}{5})}\cdot\frac{\pi}{\Gamma(\frac{4}{5})\sin\frac{\pi}{5}} \\ &= \frac{2\pi\Gamma(\frac{1}{5})}{5\sqrt{\frac{5-\sqrt{5}}{8}}\,\Gamma(\frac{2}{5})\Gamma(\frac{4}{5})} \\ &= 2\pi\sqrt{2+\frac{2}{\sqrt{5}}}\cdot\frac{\Gamma(\frac{6}{5})}{\Gamma(\frac{2}{5})\Gamma(\frac{4}{5})}.\end{aligned} \tag{8}$$

Equations (1) and (2) both hold for $0 \le t \le \pi_p/2$. We can extend the function definitions via symmetry, first to $0 \le t \le \pi_p$, then to $0 \le t < 2\pi_p$, and finally, via periodicity, to the entire real line. To wit:

$$\mathrm{sq}_p(t) = \begin{cases} \mathrm{sq}_p(\pi_p - t) & \pi_p/2 < t \le \pi_p, \\ -\,\mathrm{sq}_p(2\pi_p - t) & \pi_p < t < 2\pi_p \end{cases}$$

$$\mathrm{cq}_p(t) = \begin{cases} -\,\mathrm{cq}_p(\pi_p - t) & \pi_p/2 < t \le \pi_p, \\ \mathrm{cq}_p(2\pi_p - t) & \pi_p < t < 2\pi_p. \end{cases}$$

We finish by setting $\mathrm{sq}_p(t + 2\pi_p k) = \mathrm{sq}_p(t)$ for $k \in \mathbb{Z}$ and do the same for our cosquine function. It is not difficult to show that the squine functions are odd and the cosquine functions are even for all p. (See Figure 5.)

Squigonometric derivatives We can define the *tanquent*, *cotanquent*, *sequent*, and *cosequent* as ratios of squines and cosquines in the usual manner:

$$\mathrm{tq}_p(t) = \frac{\mathrm{sq}_p(t)}{\mathrm{cq}_p(t)}, \quad \mathrm{seq}_p(t) = \frac{1}{\mathrm{cq}_p(t)},$$

$$\mathrm{ctq}_p(t) = \frac{\mathrm{cq}_p(t)}{\mathrm{sq}_p(t)} = \frac{1}{\mathrm{tq}_p(t)}, \quad \mathrm{cseq}_p(t) = \frac{1}{\mathrm{sq}_p(t)}.$$

Equation (2) then yields two other Pythagorean-style identities:

$$\operatorname{tq}_p^p(t) + 1 = \operatorname{seq}_p^p(t), \tag{9}$$
$$1 + \operatorname{ctq}_p^p(t) = \operatorname{cseq}_p^p(t).$$

Our initial derivative formulas (1) can be combined with the usual product and quotient rules to obtain (see [**4**])

$$\frac{d}{dt}\operatorname{tq}_p(t) = \operatorname{seq}_p^2(t), \qquad \frac{d}{dt}\operatorname{seq}_p(t) = \operatorname{seq}_p^2(t)\operatorname{sq}_p^{p-1}(t), \tag{10}$$
$$\frac{d}{dt}\operatorname{ctq}_p(t) = -\operatorname{cseq}_p^2(t), \quad \frac{d}{dt}\operatorname{cseq}_p(t) = -\operatorname{cseq}_p^2(t)\operatorname{cq}_p^{p-1}(t).$$

Formulas (9) and (10) lead to the alternate form $\frac{d}{dt}\operatorname{tq}_p(t) = (1 + \operatorname{tq}_p^p(t))^{2/p}$. Separation of variables and integration yields

$$\operatorname{tq}_p^{-1}(x) = \int_0^x \frac{1}{(1+t^p)^{2/p}}\,dt, \qquad x \in \mathbb{R}.$$

The range of the arctanquent function is $(-\pi_p/2, \pi_p/2)$ [**4**]. In particular, we achieve another integral formula for π_p:

$$\pi_p = 2\int_0^\infty \frac{1}{(1+t^p)^{2/p}}\,dt.$$

A basic squigonometric integral table Exercise 10 in [**11**] challenges readers to "Use the generalized squigonometric functions to find integration formulas and identities for any $p > 1$." In [**4**], the author found several such identities. We can use these squigonometric integrals to evaluate some definite integrals whose answers were previously given only in terms of gamma or hypergeometric functions. We will also be able to come up with an expression for the volume for Hein's superegg in terms of various π_p's.

The derivative formulas listed previously can, of course, be presented as integral formulas:

$$\int \operatorname{cq}_p^{p-1}(t)\,dt = \operatorname{sq}_p(t) + C,$$
$$\int \operatorname{sq}_p^{p-1}(t)\,dt = -\operatorname{cq}_p(t) + C,$$
$$\int \operatorname{seq}_p^2(t)\,dt = \int \frac{1}{\operatorname{cq}_p^2(t)}\,dt = \operatorname{tq}_p(t) + C,$$
$$\int \operatorname{cseq}_p^2(t)\,dt = \int \frac{1}{\operatorname{sq}_p^2(t)}\,dt = -\operatorname{ctq}_p(t) + C,$$
$$\int \operatorname{seq}_p^2(t)\operatorname{sq}_p^{p-1}(t)\,dt = \int \frac{\operatorname{sq}_p^{p-1}(t)}{\operatorname{cq}_p^2(t)}\,dt = \operatorname{seq}_p(t) + C,$$
$$\int \operatorname{cseq}_p^2(t)\operatorname{cq}_p^{p-1}(t)\,dt = \int \frac{\operatorname{cq}_p^{p-1}(t)}{\operatorname{sq}_p^2(t)}\,dt = -\operatorname{cseq}_p(t) + C.$$

To develop other formulas, we use tricks we have picked up in our integration experience ([**10**], for instance), and tweak them a little.

Exercise 8 of [**11**] asks for the antiderivative of $\mathrm{sq}_4(t)\,\mathrm{cq}_4(t)$. To integrate a squine or cosquine, we need to have $p-1$ cosquines or squines in the integral in order to use substitution. We will often have to multiply and divide by powers of squines and cosquines to achieve this.

To solve Exercise 8, suppose we want to substitute $u = \mathrm{sq}_4(t)$. We will need $\mathrm{cq}_4^3(t)$ in the integral. We write

$$\int \mathrm{sq}_4(t)\,\mathrm{cq}_4(t)\,dt = \int \frac{\mathrm{sq}_4(t)\,\mathrm{cq}_4^3(t)}{\mathrm{cq}_4^2(t)}\,dt = \int \frac{\mathrm{sq}_4(t)\,\mathrm{cq}_4^3(t)}{\sqrt{1-\mathrm{sq}_4^4(t)}}\,dt.$$

We can now make our substitution $u = \mathrm{sq}_4(t)$, $du = \mathrm{cq}_4^3(t)\,dt$ to yield

$$\begin{aligned}\int \mathrm{sq}_4(t)\,\mathrm{cq}_4(t)\,dt &= \int \frac{u}{\sqrt{1-u^4}}\,du = \tfrac{1}{2}\sin^{-1}(u^2) + C\\ &= \tfrac{1}{2}\sin^{-1}(\mathrm{sq}_4^2(t)) + C.\end{aligned}$$

Double-angle identities thwarted Another source of trigonometric integral methods is the use of double-angle identities [**10**] to lower the exponent of sine and cosine expressions:

$$\cos^2(t) = \tfrac{1}{2}[1+\cos(2t)],$$
$$\sin^2(t) = \tfrac{1}{2}[1-\cos(2t)].$$

However, if we think of the argument of general squines and cosquines as an angle measure, we don't subtend angles at a constant rate if we rotate at a constant rate [**11**]. Thus we can have neither addition, subtraction, nor double-angle formulas for squigonometric functions with $p \neq 2$.

In this case, integration by parts saves the day. Suppose we want to determine $\int \mathrm{cq}_p^p(t)\,dt$. We can choose our parts as $u = \mathrm{cq}_p(t)$, $dv = \mathrm{cq}_p^{p-1}(t)\,dt$. Then $du = -\,\mathrm{sq}_p^{p-1}(t)\,dt$, $v = \mathrm{sq}_p(t)$, and we can write

$$\begin{aligned}\int \mathrm{cq}_p^p(t)\,dt &= \mathrm{cq}_p(t)\,\mathrm{sq}_p(t) + \int \mathrm{sq}_p^p(t)\,dt \qquad (11)\\ &= \mathrm{cq}_p(t)\,\mathrm{sq}_p(t) + \int [1-\mathrm{cq}_p^p(t)]\,dt\\ &= t + \mathrm{cq}_p(t)\,\mathrm{sq}_p(t) - \int \mathrm{cq}_p^p(t)\,dt.\end{aligned}$$

Solving the equation for our integral yields

$$\int \mathrm{cq}_p^p(t)\,dt = \tfrac{1}{2}[t + \mathrm{cq}_p(t)\,\mathrm{sq}_p(t)] + C. \qquad (12)$$

We can similarly derive

$$\int \mathrm{sq}_p^p(t)\,dt = \tfrac{1}{2}[t - \mathrm{cq}_p(t)\,\mathrm{sq}_p(t)] + C.$$

As a side effect of Equation (11), we also have the identity

$$\int [\mathrm{cq}_p^p(t) - \mathrm{sq}_p^p(t)]\,dt = \mathrm{cq}_p(t)\,\mathrm{sq}_p(t) + C.$$

In [**4**] there resides an extensive integral table for these functions.

Squigonometric substitution In some cases, we may be able to determine areas and volumes, or solve some indefinite integrals, using a method akin to trigonometric substitution

- If the integrand contains $a^p - x^p$, we substitute $x = a\,\mathrm{sq}_p(t)$ and use the identity $1 - \mathrm{sq}_p^p(t) = \mathrm{cq}_p^p(t)$.
- If the integrand contains $a^p + x^p$, we substitute $x = a\,\mathrm{tq}_p(t)$ and use the identity $1 + \mathrm{tq}_p^p(t) = \mathrm{seq}_p^p(t)$.
- If the integrand contains $x^p - a^p$, we substitute $x = a\,\mathrm{seq}_p(t)$ and use the identity $\mathrm{seq}_p^p(t) - 1 = \mathrm{tq}_p^p(t)$.

Suppose we wish to find the area enclosed by the hyperellipse

$$\left|\frac{x^p}{a^p}\right| + \left|\frac{y^p}{b^p}\right| = 1,$$

where $a, b > 0$. The hyperellipse is symmetric about both the x- and y-axes, so the total area A will be four times the area in the first quadrant. (See Figure 6.) We can

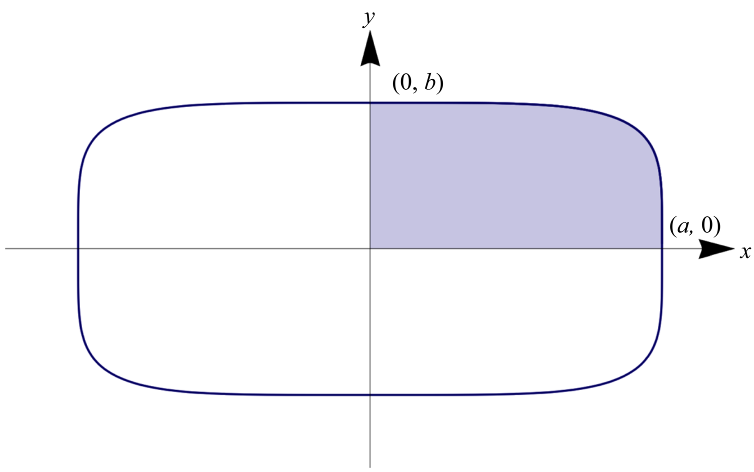


Figure 6 The hyperellipse $\left|\frac{x^p}{a^p}\right| + \left|\frac{y^p}{b^p}\right| = 1$.

solve our equation for y to get the part of the hyperellipse in the first quadrant. It is given by the function

$$y = \frac{b}{a}\sqrt[p]{a^p - x^p}, \quad 0 \le x \le a.$$

Thus

$$A = 4\int_0^a \frac{b}{a}\sqrt[p]{a^p - x^p}\,dx.$$

Set $x = a\,\mathrm{sq}_p(t)$. Then $dx = a\,\mathrm{cq}_p^{p-1}(t)\,dt$ and

$$\sqrt[p]{a^p - x^p} = \sqrt[p]{a^p - a^p\,\mathrm{sq}_p^p(t)} = \sqrt[p]{a^p\,\mathrm{cq}_p^p(t)} = a\,\mathrm{cq}_p(t),$$

since $0 \le t \le \pi_p/2$.

Then Equation (12) gives

$$\begin{aligned} A &= \frac{4b}{a}\int_0^a \sqrt[p]{a^p - x^p}\,dx \\ &= \frac{4b}{a}\int_0^{\pi_p/2} a\,\mathrm{cq}_p(t)\cdot a\,\mathrm{cq}_p^{p-1}(t)\,dt \\ &= 4ab\int_0^{\pi_p/2} \mathrm{cq}_p^p(t)\,dt \\ &= 2ab\Big[t + \mathrm{cq}_p(t)\,\mathrm{sq}_p(t)\Big]_0^{\pi_p/2} \\ &= \pi_p ab. \end{aligned}$$

Via Google Earth measurements, we have for Sergels Torg that $a \approx 55$ m and $b \approx 66$ m (measuring from outer curb to outer curb), giving an area for the Stockholm superellipse of $\pi_{2.5}ab \approx 3.38094(55)(66) \approx 12{,}273$ m^2, or about 3 acres.

We get two side benefits from the previous calculation. On the one hand, if we were to back substitute for t at the end of the example, we get the antiderivative formula

$$\int \sqrt[p]{a^p - x^p}\,dx = \frac{1}{2}\left[a^2\,\mathrm{sq}_p^{-1}\left(\frac{x}{a}\right) + x\sqrt[p]{a^p - x^p}\right] + C.$$

On the other, if we set $a = b = 1$, we calculate the area of the unit squircle to be π_p and gain a third integral formula for π_p:

$$\pi_p = 4\int_0^1 \sqrt[p]{1 - x^p}\,dx.$$

Volume of the superegg From the previous calculation, we can find an expression for the volume of Hein's superegg:

$$\begin{aligned} V &= 2\int_0^a \pi\left(\frac{b}{a}\sqrt[p]{a^p - x^p}\right)^2 dx \\ &= \frac{2\pi b^2}{a^2}\int_0^a (a^p - x^p)^{2/p}\,dx \\ &= \frac{2\pi b^2}{a^2}\int_0^{\pi_p/2} [a\,\mathrm{cq}_p(t)]^2 \cdot a\,\mathrm{cq}_p^{p-1}(t)\,dt \\ &= 2\pi ab^2\int_0^{\pi_p/2} \mathrm{cq}_p^{p+1}(t)\,dt. \end{aligned}$$

Integration by parts yields the reduction formula ([**4**, Formula (42)])

$$\int \mathrm{cq}_p^m(t)\,dt = \frac{1}{m-p+2}\,\mathrm{cq}_p^{m-p+1}(t)\,\mathrm{sq}_p(t) + \frac{m-p+1}{m-p+2}\int \mathrm{cq}_p^{m-p}(t)\,dt$$

for $m \neq p-2$. So with $m = p+1$ we have

$$V = 2\pi ab^2\left[\frac{1}{3}\,\mathrm{cq}_p^2(t)\,\mathrm{sq}_p(t)\Big|_0^{\pi_p/2} + \frac{2}{3}\int_0^{\pi_p/2} \mathrm{cq}_p(t)\,dt\right]$$

$$= \frac{4\pi ab^2}{3} \int_0^{\pi_p/2} \mathrm{cq}_p(t)\, dt.$$

We multiply and divide the integrand by $\mathrm{cq}_p^{p-2}(t)$ to get

$$\begin{aligned} V &= \frac{4\pi ab^2}{3} \int_0^{\pi_p/2} \frac{\mathrm{cq}_p^{p-1}(t)}{\mathrm{cq}_p^{p-2}(t)}\, dt \\ &= \frac{4\pi ab^2}{3} \int_0^{\pi_p/2} \frac{\mathrm{cq}_p^{p-1}(t)}{(1-\mathrm{sq}_p^p(t))^{(p-2)/p}}\, dt. \end{aligned}$$

Substituting $u = \mathrm{sq}_p^p(t)$ gives $u^{1/p} = \mathrm{sq}_p(t)$, $\frac{1}{p} u^{1/p-1}\, du = \mathrm{cq}_p^{p-1}(t)\, dt$ and

$$\begin{aligned} V &= \frac{4\pi ab^2}{3p} \int_0^1 u^{1/p-1}(1-u)^{2/p-1}\, du \\ &= \frac{4\pi ab^2}{3p} \mathrm{B}(\tfrac{1}{p}, \tfrac{2}{p}) \\ &= \frac{4\pi ab^2}{3p} \frac{\Gamma(\frac{1}{p})\Gamma(\frac{2}{p})}{\Gamma(\frac{3}{p})}. \end{aligned}$$

We can now determine exact values for the volumes of some supereggs by use of the identities (6). When $p = 1$ we obtain $V = \frac{2\pi ab^2}{3} \approx 2.0944ab^2$, rediscovering the formula for the volume of a (double) cone in the process.

For $p = 2$, $V = \frac{4\pi ab^2}{3} \approx 4.1888ab^2$. When $p = 3$, equation (7) yields

$$V = \frac{4\pi ab^2}{9} \cdot \Gamma(\tfrac{1}{3})\Gamma(\tfrac{2}{3}) = \frac{4\pi ab^2}{9} \cdot \frac{2\pi}{\sqrt{3}} = \frac{8\pi^2 ab^2}{3^{5/2}} \approx 5.0651ab^2.$$

For $p = 4$, equation (7) gives

$$\begin{aligned} V &= \frac{4\pi ab^2}{12} \cdot \frac{\Gamma(\frac{1}{4})\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{4})} \\ &= \frac{\pi ab^2}{3} \frac{[\Gamma(\frac{1}{4})]^2 \cdot \sqrt{\pi}}{\sqrt{2}\pi} \\ &= \frac{\pi ab^2}{3} \frac{2\pi_4\sqrt{\pi} \cdot \sqrt{\pi}}{\sqrt{2}\pi} \\ &= \frac{\pi_4 \pi \sqrt{2} ab^2}{3} \approx 5.4916ab^2. \end{aligned}$$

Finally, we can use equation (8) to get the volume of Piet Hein's superegg with $p = 2.5$:

$$V = \frac{8\pi ab^2}{15} \cdot \frac{\Gamma(\frac{2}{5})\Gamma(\frac{4}{5})}{\Gamma(\frac{6}{5})} = \frac{16\sqrt{2 + \frac{2}{\sqrt{5}}}\,\pi^2 ab^2}{15\pi_5} \approx 4.7126ab^2.$$

In particular, the "antistress" superegg available online has height $a = 3.5$ cm and width $b = 2.5$ cm [**8**], and thus its volume is approximately 103.087 cm^3.

REFERENCES

1. C. L. Adler, J. Tanton, π is the minimum value of Pi, *College Math. J.* **31** (2000) 102–106.
2. R. Euler, J. Sadek, The pi's go full circle, *Math. Mag.* **72** (1999) 59–63.
3. M. Gardner, *The Colossal Book of Mathematics*. W.W. Norton & Company, New York, 2001.
4. R. Grammel, Eine Verallgemeinerung der Kreis- und Hyperbelfunktionen, *(German) Ing.-Arch.* **16** (1948) 188–200.
5. Piet Hein—Classic Danish Design, http://www.piethein.com.
6. J. Lang, D. Edmunds, *Eigenvalues, Embeddings and Generalised Trigonometric Functions*. Lecture Notes in Mathematics, Vol. 2016. Springer-Verlag, Berlin, 2011.
7. B. MacLennan, L_p-circular functions (1993), http://web.eecs.utk.edu/~mclennan/papers/LpCirc-Jan93.pdf.
8. http://www.panik-design.com/acatalog/Piet_Hein___Super_Egg.html.
9. D. Shelupsky, A generalization of the trigonometric functions, *Amer. Math. Monthly* **66** (1959) 879–884.
10. J. Stewart, *Calculus, Early Transcendentals*. Sixth edition. Thomson Brooks/Cole, Belmont, CA, 2008.
11. W. Wood, Squigonometry, *Math. Mag.* **84** (2011) 257–265.

Summary. A superegg is the solid of revolution for a hyperellipse, an ellipse with squarish corners. The two objects became the basis of a design revolution of sorts in the 1960s, as practiced by the Danish mathematician and poet Piet Hein. We use an analog of trigonometry called squigonometry to produce a set of constants akin to the well-known and customary π and then find formulas using these constants for the area of a hyperellipse and the volume of Hein's superegg.

ROB POODIACK (MR Author ID: 886276) received his Ph.D. in mathematical sciences in 1999 from the University of Vermont. He is currently Professor and Chair of the Department of Mathematics at Norwich University in Northfield, VT, the nation's oldest private military college. His interests include his two teenaged sons, finding novel ways to evaluate integrals and series, and continual training for the next half-marathon.

1 S	2 T	3 A	4 M	5 P	■	6 C	7 A	8 T	■	9 R	10 O	11 T	12 O	13 R
14 H	E	M	A	L	■	15 O	N	E	■	16 U	N	I	T	Y
17 E	M	I	L	E	■	18 M	T	S	■	19 S	I	T	I	N
■	20 P	R	O	B	21 A	B	I	L	22 I	S	T	I	C	■
■	■	■	23 N	E	B	■	■	24 A	N	I	■	■	■	■
25 A	26 R	27 S	E	■	28 A	29 R	30 K	■	31 J	A	32 C	33 O	34 B	35 I
36 L	E	O	■	■	37 C	A	L	38 I	■	39 N	O	D	E	S
40 I	N	F	41 I	42 N	I	T	E	D	43 E	S	C	E	N	T
44 G	E	E	S	E	■	45 T	I	E	R	■	■	46 T	I	O
47 N	E	W	T	O	48 N	■	49 N	S	A	■	50 M	O	N	O
■	■	■	■	51 L	E	52 D	■	■	53 S	54 O	Y	■	■	■
■	55 F	56 I	57 N	I	T	E	58 E	59 L	E	M	E	60 N	61 T	■
62 V	O	M	I	T	■	63 U	N	I	■	64 E	L	I	O	65 T
66 L	U	R	C	H	■	67 C	O	N	■	68 G	I	N	A	S
69 T	R	E	E	S	■	70 E	S	E	■	71 A	N	O	D	E

Proof Without Words: Infinitely Many Almost-Isosceles Pythagorean Triples Exist

ROGER B. NELSEN
Lewis & Clark College
Portland, OR
nelsen@lclark.edu

An *almost-isosceles Pythagorean triple* is a triple $(a, a+1, c)$ of positive integers such that $a^2 + (a+1)^2 = c^2$. We prove the existence of infinitely many such triples via the following lemma, proved using the inclusion–exclusion principle.

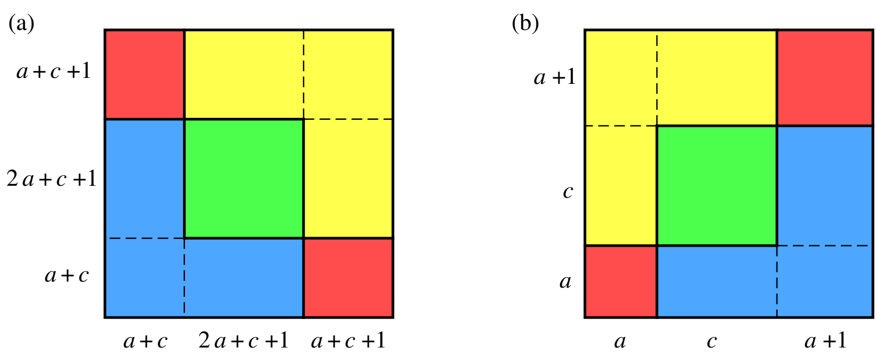
Lemma.
(a)

$$(4a+3c+2)^2 = (3a+2c+1)^2 + (3a+2c+2)^2 \Leftrightarrow$$
$$(2a+c+1)^2 = 2(a+c)(a+c+1).$$

(b)

$$(2a+c+1)^2 = 2(a+c)(a+c+1)^2 \Leftrightarrow$$
$$a^2 + (a+1)^2 = c^2.$$

Proof.



(a)

$$(4a+3c+2)^2 =$$
$$(3a+2c+1)^2 + (3a+2c+2)^2 - (2a+c+1)^2 + 2(a+c)(a+c+1).$$

(b)

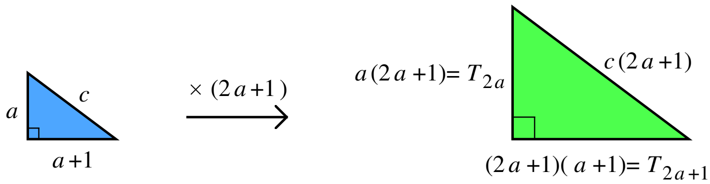
$$(2a+c+1)^2 = 2(a+c)(a+c+1) - c^2 + a^2 + (a+1)^2.$$

Math. Mag. **89** (2016) 103–104. doi:10.4169/math.mag.89.2.103.
MSC: Primary 05A15; Secondary 11A25

Theorem. *Infinitely many almost-isosceles Pythagorean triples exist since* $3^2+4^2=5^2$ *and* $a^2+(a+1)^2=c^2 \Leftrightarrow (3a+2c+1)^2+(3a+2c+2)^2=(4a+3c+2)^2$. *E.g.* $3^2+4^2=5^2 \Leftrightarrow 20^2+21^2=29^2 \Leftrightarrow 119^2+120^2=169^2$, etc.

Corollary. *There exist infinitely many Pythagorean triples in which the legs are consecutive triangular numbers.*

Proof (where $T_n=1+2+\cdots+n=n(n+1)/2$ is the n^{th} triangular number).



E.g. $3^2+4^2=5^2 \Leftrightarrow T_6^2+T_7^2=(7\cdot 5)^2$, $20^2+21^2=29^2 \Leftrightarrow T_{40}^2+T_{41}^2=(41\cdot 29)^2$, $119^2+120^2=169^2 \Rightarrow T_{238}^2+T_{239}^2=(239\cdot 169)^2$, etc.

Summary. Wordlessly, we show that there are infinitely many Pythagorean triples with consecutive integers as legs and infinitely many Pythagorean triples with consecutive triangular numbers as legs.

REFERENCE

1. W. Sierpiński, Sur les nombres triangulaires carrés, *Bull. Soc. Roy. Sci. Liège*, **30** (1961) 189–194.

ROGER NELSEN (MR Author ID: 237909) is professor emeritus at Lewis & Clark College, where he taught mathematics and statistics for 40 years.

Applying “New Math”

In the fall of 1964, our Algebra II class started off with six weeks of “new math” (set theory), as mandated by the Louisiana State Board of Education. We found a way to solve some algebra problems using set theory when our next door neighbor asked for help with simultaneous equations involving two-digit numbers.

Example. A two-digit number has ten’s digit four more than the unit’s digit and the sum of the two digits is eight. What is the number?

Solution. Let S be the set of two-digit numbers whose ten’s digit is four more than the unit’s digit. Then, $S=\{40, 51, 62, 73, 84, 95\}$. Let T be the set of two-digit numbers whose ten’s digits sum to 8. So, $T=\{17, 26, 35, 44, 53, 62, 71, 80\}$. The answer to the question is 62 because $S\cap T=\{62\}$.

Using set theory to solve such a problem when teaching future elementary math teachers can be useful because it shows how a seemingly unrelated algebra problem can be solved using set theory.

— Howard Cary Morris
Southwest Tennessee Community College

Spacewalks and Amusement Rides: Illustrations of Geometric Phase

JEFFREY LAWSON
Department of Mathematics and Computer Science
Western Carolina University
Cullowhee, North Carolina 28723
jlawson@wcu.edu

MATTHEW RAVE
Department of Chemistry and Physics
Western Carolina University
Cullowhee, North Carolina 28723
mrave@wcu.edu

The Sidewinder at Six Flags Over Texas is a spinning amusement park ride consisting of three clusters of four pods. Each cluster is connected to a center rotating arm and each cluster pivots on its own axis, creating a "spin within a spin." Let's suppose you

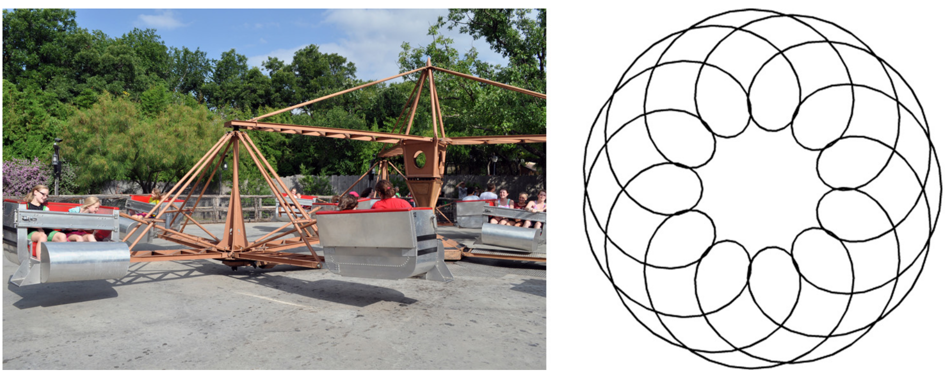


Figure 1 Sidewinder (reproduced with permission, copyright 2010 GuideToSFoT.com) and its trajectory.

jump onto the ride and your friend stays at the gate to take a photo of you looking very dizzy. How often does your friend have an opportunity for that great photo? Think about it: not only should your cluster have completed a full circle but also your pod needs to be at its maximum distance from the center hub of the ride. If, say, the clusters revolve around the central axis at 9 rpm and the pods rotate around a cluster axis at 24 rpm, then how many revolutions will it take between photo ops? Since the clusters rotate $8/3$ faster than the main contraption then for every three revolutions of the whole ride around the central axis, the pods will rotate around a cluster eight complete times. Thus your friend only has to wait for the ride to go around three times before taking another picture.

Although the calculation is simple, is there a systematic way to address such a question? The answer involves the concept of *geometric phase*—a concept with roots in physics but which is entirely mathematical in nature. The geometric phase will help

Math. Mag. **89** (2016) 105–121. doi:10.4169/math.mag.89.2.105.
MSC: Primary 37J15, Secondary 53D20; 70H33

us identify what amusement park rides, a Chopin étude, Spirographs, and quantum mechanics all have in common.

In 1984, Michael Berry published a paper [**5**] on the subject of *phase*, spawning an entirely new branch of physics. Indeed, Berry's phase was dubbed "the phase that launched 1000 scripts" [**24**] because of the number of research avenues it opened. Even the classical mechanical analogue (also called Hannay's phase [**9**]) led to important insights, such as an intuitive approach to the Foucault pendulum [**4, 10, 20**], a device so often displayed in science museums. The ubiquity of Berry's phase in quantum mechanics has made it a popular topic for books [**8, 19, 23**] and review articles [**2, 6, 22**] in physics. Classical applications are rarer in the literature and are for the most part limited to more advanced treatises on mechanics [**7, 11, 15, 16, 17**]. On the other hand, treatments in the undergraduate mathematical literature have been almost nonexistent, despite the fact that Berry's phase and its classical analogue (collectively, geometric phase) are, at their core, manifestations of simple mathematical properties of a system.

We seek to partially fill this void by describing how geometric phase arises as a specific kind of phase that occurs in a wide variety of problems. We will investigate the underlying mathematics of geometric phase and include some illuminating examples. Geometric phase gives us a glimpse into ***geometric mechanics***, the use of geometry to solve mechanics problems.

Phase difference and geometric phase

Clocks. Consider a clock with a minute hand and an hour hand. When the minute hand completes one cycle, the hour hand has completed one-twelfth of a cycle. Thus when the minute hand goes around 2π radians, the hour hand subtends an angle of $\frac{1}{12}(2\pi) = \frac{\pi}{6}$ radians. It is natural to think of this angle as the phase difference between the two hands.

More generally, for two periodic functions f_1 and f_2 with periods T_1 and T_2, respectively, the ***phase difference*** γ between f_1 and f_2 is the radian value that f_2 subtends if f_1 completes exactly one cycle, that is,

$$\gamma = 2\pi \left(\frac{T_1}{T_2} \right). \tag{1}$$

We have assumed, without loss of generality, that $T_1 < T_2$ so that $0 \leq \gamma \leq 2\pi$.

The key observation about the clock hands is that T_1 and T_2 are dependent. That is, it is not possible to wind one hand of a clock without affecting the other hand. This leads us to the concept of geometric phase. A phase difference such as (1) is said to be ***geometric*** if it depends solely upon a nontrivial relationship between the variables in question and not upon the rate at which the cycles are traversed. (If a phase difference depends explicitly upon time, it is said to be ***dynamic***.) Crucially, if T_2 changes then T_1 will change too, so that γ (which depends upon a *ratio* of periods) remains the same. Such a connection is the hallmark of a geometric phase.

Pendulums. Consider a pair of pendulums of lengths ℓ_1 and ℓ_2 situated in the same laboratory. The phase difference between them does not seem to be geometric because the periods are not obviously linked. Indeed, the periods T_1 and T_2 depend only on the lengths, and if you imagine cutting one pendulum in half, the other pendulum would be unchanged. However, suppose *the laboratory itself* could be moved and taken, say, to the moon. We now have a different system entirely: The periods are now functions of *two* variables, ℓ and g. The acceleration due to gravity g is no longer fixed, and g

becomes a parameter that connects the two pendulums. Since the period of a pendulum in a gravitational field is given by $T = 2\pi\sqrt{\ell/g}$ for small oscillations, we find

$$\frac{T_1}{T_2} = \sqrt{\frac{\ell_1}{\ell_2}}.$$

Since the ratio of the periods is fixed, the phase difference is clearly geometric. Placing both pendulums on the moon would change both periods but their ratio would remain constant.

Polyrhythms. In music, when two rhythms are not multiples of each other (polyrhythm), the underlying "global" rhythm is greater than each of the components. This global rhythm is based upon the least common multiple of the two periods. In Chopin's Trois nouvelles études, No. 1 (Figure 2), the left hand plays eight notes for every six in the right hand.

Figure 2 A polyrhythm from the 32nd measure of Chopin's Trois nouvelles études, No. 1.

Let $T_1 = 1/8$ be the duration (period) of a "left-hand" note (in this case, notated by Chopin with an eighth note), and let $T_2 = 1/6$ be the duration of a "right-hand" quarter note. Then $\gamma = 2\pi\left(\frac{1/8}{1/6}\right) = 2\pi\ (6/8) = 3\pi/2$. This represents how far the right hand gets through its rhythmic "cycle" with each note played in the left hand. To see that this is indeed a geometric phase, consider that there is a parameter that connects the left hand to the right, namely the metronome marking M of the piece: the tempo or "rate" at which it is to be played. The periods T_1 and T_2 are clearly *dependent* on M, but the phase difference is *geometric* because the *ratio* T_1/T_2 is not. If the pianist plays the left-hand part faster or slower, the right-hand part must follow suit—otherwise the pianist is not playing the music as written.

Spirographs. Spirographs are drawn with one circle rotating around the inside of a larger circle. Mathematically, the circles trace what are called ***hypotrochoids***. We can relate hypotrochoids to the geometric phase by noting that the smaller circle (radius r) has a different period than the larger circle (radius R), *no matter how quickly or how slowly the circles are drawn*. The general formula for a hypotrochoid is

$$x(t) = (R - r)\cos t + h\cos\frac{(R - r)t}{r}, \tag{2}$$

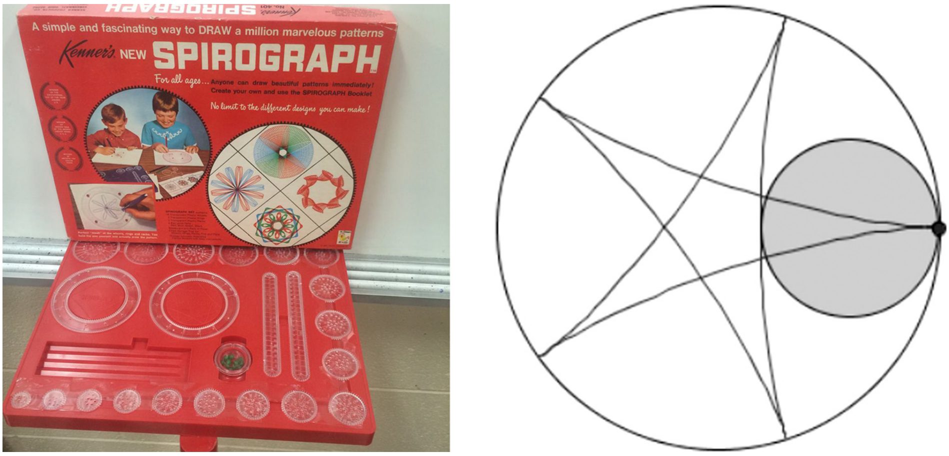


Figure 3 Hasbro Spirograph, circa 1967 and a hypotrochoid with $R/r = 5/2$ and offset $h = r$. You can generate your own at [**14**].

$$y(t) = (R - r)\sin t - h \sin \frac{(R-r)t}{r}.$$

Here h is the "pen offset," namely, the distance from the center of the smaller rolling circle to the point where the curve is traced. We will henceforth take $h = r$, so that the "pen" is on the outside edge of the small circle. The shape of each Spirograph depends on the ratio R/r. Assume that R/r is rational with reduced form p/q. (For example, if $R = 120$ and $r = 48$ then $p/q = 5/2$.) Then,

1. The numerator p is the number of cusps in the figure (i.e., the number of times the pen touches the larger circle).
2. The denominator q is the number of times the smaller circle goes around inside the big circle until the pen returns to its original position.

The geometric phase of the system is $\gamma = ((R - r)/r)\,2\pi$. This can be seen intuitively by noting that two frequencies appear in the parameterizations of x and y in (2)—that is, $\omega_1 = 1$ and $\omega_2 = (R - r)/r$. If $\gamma < 2\pi$, the period traversing the big circle is smaller than the period of the little circle's rotation. If $\gamma > 2\pi$, then the little circle rotates through *more* than a full revolution for every traversal of the big circle.

Quantum mechanics. As we mentioned previously, Berry's phase originated in quantum mechanics. In quantum mechanics, particles exist in states $\psi(\mathbf{r}, t)$ which are solutions to the complex eigenvalue equation known as Schrödinger's equation

$$\mathcal{H}(t)\psi(\mathbf{r}, t) = E(t)\psi(\mathbf{r}, t).$$

Here $\mathcal{H}(t)$ is an operator and $E(t)$ is an energy eigenvalue. Since such solutions are only unique *up to an arbitrary phase* $e^{i\phi}$, any states $\psi(\mathbf{r}, t)$ and $e^{i\phi}\psi(\mathbf{r}, t)$ should be physically indistinguishable, in the sense that measurements depend upon the modulus squared $|e^{i\phi}\psi(\mathbf{r}, t)|^2 = |\psi(\mathbf{r}, t)|^2$ rather than the state itself. However, Berry's work [**5**] in 1984 changed this view. Berry discovered that the phase difference $\gamma = \phi_2 - \phi_1$ between two ostensibly similar states—until then a quantity assumed to have no physical relevance—is actually measurable in certain situations. For example, in the Aharonov–Bohm effect [**1**], the presence of an electromagnetic vector potential $\mathbf{A}$ is seen to have physical implications even though the consequent magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ is zero.

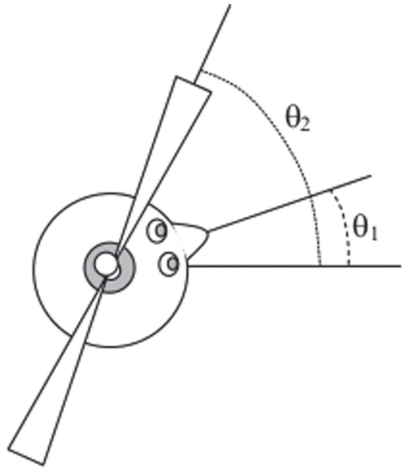


Figure 4 Elroy wearing a beanie cap with a propeller atop.

Calculating geometric phase

So how do we calculate geometric phase in mechanical systems? Every example so far is a system with more than one period, in which the periods "synchronize" at regular intervals. This suggests that geometric phase can always be defined for a mechanical system with least two periods—as long as the periods are dependent in some way. Therefore, any system whose equations of motion represent a closed orbit should be characterizable by a geometric phase.

In our examples thus far, we found the periods T_1 and T_2 by computing the orbits and then calculating geometric phase directly using (1). In general this "brute force" method requires finding an explicit solution to a second order initial value problem. However, we may invoke a geometric method. Namely, by using a symmetry argument and/or a physical conservation law, we can derive a differential form that relates all of the generalized coordinates. It is then possible to find the geometric phase by integrating this differential form. The complete equations of motion (EOM) are not necessary.

Elroy's beanie. Elroy's beanie [**15, 16**] depicts a boy Elroy floating in space, with only the propeller on his beanie hat to control his orientation (Figure 4). The system consists of two planar rigid bodies rotating about a common pin joint that pierces each rigid body through its center of mass. We declare an inertial xy-coordinate frame for this system and assign the first rigid body (Elroy floating in space) moment of inertia I_1 and angle θ_1 as measured counterclockwise from the x-axis. The orientation moment of inertia I_2 and θ_2 of the second body (the propeller) are measured likewise. The energy for this system with no external forces (that is, the *free* system) is the kinetic energy ***Lagrangian***

$$L(\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2) = \frac{1}{2}\left(I_1\dot{\theta}_1^2 + I_2\dot{\theta}_2^2\right), \qquad (3)$$

where a "dot" denotes a time derivative, $\dot{q} = \frac{dq}{dt}$.

Intuitively, the only thing that matters physically is the angle between the major axes of the two rigid bodies. This motivates a change to a new set of coordinates

$$\theta = \theta_1 \quad \text{and} \quad \psi = \theta_2 - \theta_1.$$

Note that θ is physically arbitrary, merely representing the orientation of the system with respect to the fixed laboratory frame. The difference coordinate ψ, on the other hand, contains all of the essential dynamics of the system. In these new coordinates, (3) becomes

$$L(\theta, \psi, \dot{\theta}, \dot{\psi}) = \frac{1}{2}\left(I_1\dot{\theta}^2 + I_2(\dot{\theta} + \dot{\psi})^2\right). \tag{4}$$

To determine the equations of motion, the ***generalized momenta***,

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = I_1\dot{\theta} + I_2(\dot{\theta} + \dot{\psi}) \quad \text{and} \quad p_\psi = \frac{\partial L}{\partial \dot{\psi}} = I_2(\dot{\theta} + \dot{\psi}), \tag{5}$$

are substituted into the ***Euler–Lagrange equations***

$$\dot{p}_\theta - \frac{\partial L}{\partial \theta} = 0 \quad \text{and} \quad \dot{p}_\psi - \frac{\partial L}{\partial \psi} = 0, \tag{6}$$

which describe the evolution of the physical system. (In mechanics, the Euler–Lagrange equations generalize Newton's second law of motion. They can be derived using Hamilton's principle of stationary action. See, for example, [**3**]) These equations reduce to the trivial system of equations $\ddot{\theta} = 0 = \ddot{\psi}$, and the solutions are the equations of motion

$$\theta(t) = \omega_\theta\, t \quad \text{and} \quad \psi(t) = \omega_\psi\, t, \tag{7}$$

where we have assumed the initial conditions $\theta(0) = 0$, $\psi(0) = 0$, $\dot{\theta}(0) = \omega_\theta$, and $\dot{\psi}(0) = \omega_\psi$. Note that neither angle variable appears explicitly in (4), which, by applying (5) and (6), implies that their corresponding momenta p_θ and p_ψ are constant. In retrospect, (7) could have been written down by inspection.

Because there are no external influences, the total angular momentum p_θ is constant in this system. If the system as a whole does not drift rotationally, we may set $p_\theta = 0$. Then (5) and (7) relate ω_θ and ω_ψ by

$$\frac{\omega_\theta}{\omega_\psi} = \frac{-I_2}{I_1 + I_2}. \tag{8}$$

It is clear from (8) that $|\omega_\theta| \leq |\omega_\psi|$, which implies that their periods satisfy $|T_\theta| \geq |T_\psi|$. Combining (1) and (8), our "brute force" method reveals the geometric phase to be

$$\gamma = 2\pi\left(\frac{T_\psi}{T_\theta}\right) = 2\pi\left(\frac{\omega_\theta}{\omega_\psi}\right) = -2\pi\left(\frac{I_2}{I_1 + I_2}\right).$$

For example, if $I_1 = I_2$, then $\gamma = -\pi$, which means that one complete spin of the propeller (relative to Elroy) turns Elroy a half turn in the opposite direction (relative to the lab). This corresponds to $\dot{\theta}_1 = -\dot{\theta}_2$ in the original coordinates. Note that, regardless of which moment of inertia is larger, $|\gamma| < 2\pi$.

What's so special about p_θ? After all, there are several independent constants of the motion (i.e., conserved quantities), including L and p_ψ. The answer lies in a symmetry argument, which will provide us our first glimpse behind the scene when we calculate geometric phase.

All possible states of our system can be specified by giving the set of all possible configurations (θ, ψ) on the ***configuration space*** M of the system. Here M is the torus $T^2 = S^1 \times S^1$. To describe the motion, it suffices to find an expression for $\psi(t)$.

Recall that θ is physically arbitrary—because different values of θ represent the same dynamics regardless of the point of reference, we say that θ is a symmetry variable. In other words, only one of the two copies of the circle S^1 in $T^2 = S^1 \times S^1$ is "important" physically, and the other represents a circular symmetry of the mechanical system. We say that the ***symmetry group*** is the real special orthogonal group $SO(2)$, represented here by S^1 under addition of angles. This tells us that rotating the entire configuration by a fixed angle α, say, has no effect upon the dynamics.

Given a configuration space M, the ***action*** of G on M is a mapping of G to the space of bijective differential mappings of M to itself (*diffeomorphisms* of M), in which multiplication in the group maps to composition of operators. In this case, we have $SO(2)$ acting on T^2, rotating the system by an angle α. This rotation only affects θ: we write

$$\alpha \cdot (\theta, \psi) = (\alpha + \theta, \psi), \tag{9}$$

from which we can verify that

$$\alpha \cdot (\beta \cdot (\theta, \psi)) = (\alpha + \beta + \theta, \psi).$$

For comparison, in the original coordinate system the action is

$$\alpha \cdot (\theta_1, \theta_2) = (\alpha + \theta_1, \alpha + \theta_2).$$

Although the constants of the motion L, p_1, p_2, p_θ, and p_ψ are all invariant under this action, the new set of coordinates concisely describes our observations that the "true" dynamics lie only in the relative angle ψ between the rigid bodies. We have effectively used a symmetry to simplify the dynamical system. Our motivation to transform from coordinates $\{\theta_1, \theta_2\}$ to coordinates $\{\theta, \psi\}$ is to isolate θ, the only coordinate affected by the group action. The total angular momentum p_θ is special because it is the momentum corresponding to the variable affected by the action of the symmetry group in (9). Thus we find the geometric phase by starting with conservation of momentum in the θ-direction.

Unfortunately, in many systems the equations of motion are difficult if not impossible to derive analytically. In such cases, the geometric phase may often still be found using a geometric argument. For example, in Elroy's beanie, the total angular momentum (which is conserved) is given by $p_\theta = \mu$, where $\mu \in \mathbb{R}$. We can use the first equation of (5) to construct the differential form $d\theta$ as follows:

$$\mu = (I_1 + I_2)\dot{\theta} + I_2\dot{\psi},$$
$$\mu\, dt = (I_1 + I_2)d\theta + I_2 d\psi. \tag{10}$$

If we restrict ourselves to the zero-angular-momentum solution, then we have

$$0 = d\theta + \frac{I_2}{I_1 + I_2} d\psi, \tag{11}$$

which is exactly the relationship between variables that produces a geometric phase. Integrating around one closed circuit C in the angle ψ gives us the geometric phase

$$\gamma = \Delta\theta = \int_C d\theta = \int_0^{2\pi} -\frac{I_2}{I_1 + I_2}\, d\psi = -2\pi \left(\frac{I_2}{I_1 + I_2} \right).$$

For the free beanie finding the equations of motion is simple, so the geometric method has no particular advantage. It is more advantageous in the case of a beanie with a spring-loaded propeller. The spring force may be modeled by subtracting a potential energy term $V(\psi)$ from the kinetic energy Lagrangian in (3), giving a new Lagrangian

$$\tilde{L}(\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2) = \frac{1}{2}\left(I_1\dot{\theta}_1^2 + I_2\dot{\theta}_2^2\right) - V(\psi).$$

Applying (5) to $\tilde{L}$ does not change the formulas for momenta p_θ and p_ψ, but p_ψ is no longer a constant of the motion. The Lagrangian $\tilde{L}$ and the total angular momentum p_θ are invariant under the action of $SO(2)$, so we may ignore the potential energy in computing geometric phase.

Elroy's fashionable hat. Now what if for Elroy's beanie we allow one of the rigid bodies to have a variable moment of inertia? Again, assign Elroy a constant moment

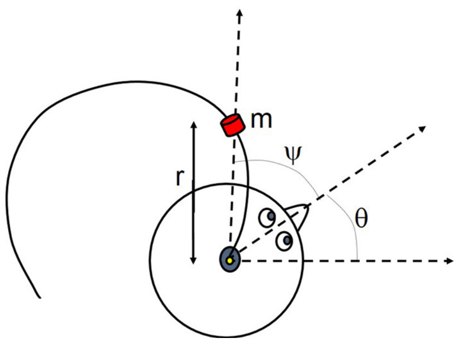


Figure 5 Elroy's fashionable hat.

of inertia I_1 and measure angle θ counterclockwise from the x-axis. Assume that the propeller is not straight as before, but instead is a stiff massless wire spiraling outward (Figure 5). A point mass m can be adjusted to any position along the wire, and will slide along the wire as the propeller spins. The angle ψ is now the angle that Elroy makes with the ray that extends from the origin to the mass m. As the mass slides along the propeller, ψ changes, and the distance r between the origin and and the mass follows the spiral equation of the wire

$$r = \sqrt{\frac{c\psi}{m}},$$

where c is a positive constant with units of moment of inertia. We picked this equation for the wire ad hoc, in order to force the moment of inertia to be linear in ψ. That is, a propeller with this exact shape has moment of inertia I_2 given by

$$I_2(\psi) = mr^2 = c\psi.$$

The energy for this system with no external forces (that is, the *free* system) is given by the kinetic energy Lagrangian

$$L(\theta, \psi, \dot{\theta}, \dot{\psi}) = \frac{1}{2}\left(I_1\dot{\theta}^2 + c\psi(\dot{\theta} + \dot{\psi})^2\right).$$

Observe that θ still does not appear explicitly in the Lagrangian (but ψ does), so that the generalized momentum

$$\mu = p_\theta = \frac{\partial L}{\partial \dot{\theta}} = I_1\dot{\theta} + c\psi(\dot{\theta} + \dot{\psi})$$

is a conserved quantity. After letting $\mu = 0$ again, we can readily write the equation of differential forms

$$0 = d\theta + \frac{c\psi}{I_1 + c\psi} d\psi \tag{12}$$

that relates changes in θ to changes in ψ in the case where the total angular momentum μ is zero. In general, for any μ, and for $0 \le \psi \le 2\pi$, the change in angle θ is then given by

$$\Delta\theta = \Delta\theta_{\text{dynamic}} + \Delta\theta_{\text{geometric}},$$

where

$$\Delta\theta_{\text{dynamic}} = \int_C \frac{\mu}{I_1 + c\psi(t)} dt$$

and

$$\Delta\theta_{\text{geometric}} = -\int_0^{2\pi} \frac{c\psi}{I_1 + c\psi} d\psi = \frac{I_1}{c} \ln\left(\frac{I_1 + 2\pi c}{I_1}\right) - 2\pi.$$

If $c = I_1/2$, then $\Delta\theta_{\text{geometric}} \approx -1.1\pi$. Although we need to find $\psi(t)$ explicitly to compute the dynamic phase, again we need not solve the differential equations to compute the geometric phase.

Annular amusement ride. Our next example is a system in which a massless rod of length r_1 is free to rotate about the origin. At the end of the rod is a hinge of mass m_1, to which is attached another massless rod of length $r_2 \le r_1$. At the end of this rod is another mass m_2. We call this system the "annular amusement ride" because the outer point mass stays within the open annulus $(r_1 - r_2) < r < (r_1 + r_2)$, and its motion models spinning amusement park rides such as the aforementioned Sidewinder.

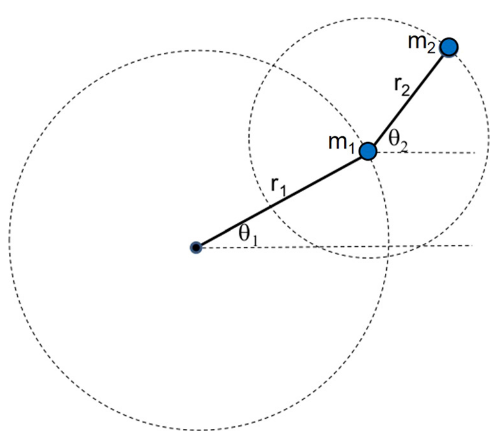


Figure 6 The annular amusement ride.

A natural choice of coordinate system is $\{\theta_1, \theta_2\}$, where the angles are measured with respect to the fixed laboratory frame (see Figure 6). The Cartesian coordinates of the inner mass m_1 are

$$x_1 = r_1 \cos\theta_1 \quad \text{and} \quad y_1 = r_1 \sin\theta_1,$$

whereas the coordinates of the outer mass m_2 are

$$x_2 = r_1 \cos\theta_1 + r_2 \cos\theta_2 \quad \text{and} \quad y_2 = r_1 \sin\theta_1 + r_2 \sin\theta_2.$$

From this we can write the free Lagrangian

$$L(x, y) = \frac{1}{2} m_1 \left(\dot{x_1}^2 + \dot{y_1}^2\right) + \frac{1}{2} m_2 \left(\dot{x_2}^2 + \dot{y_2}^2\right)$$

as

$$L(\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2) = \frac{1}{2}(m_1 + m_2) r_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 {r_2}^2 \dot{\theta}_2^2 + m_2 r_1 r_2 \dot{\theta}_1 \dot{\theta}_2 \cos(\theta_2 - \theta_1).$$

As before, switch to coordinates $\{\theta, \psi\}$ defined by $\theta = \theta_1$ and $\psi = \theta_2 - \theta_1$. In these coordinates, θ represents the orientation of the inner rod with respect to a fixed laboratory frame, whereas ψ represents the relative angle between the outer rod and an extension of the inner rod. The Lagrangian becomes

$$L(\theta, \psi, \dot{\theta}, \dot{\psi}) = \frac{1}{2}(m_1 + m_2) {r_1}^2 \dot{\theta}^2 + \frac{1}{2} m_2 {r_2}^2 (\dot{\theta} + \dot{\psi})^2 + m_2 r_1 r_2 \dot{\theta}(\dot{\theta} + \dot{\psi}) \cos\psi. \tag{13}$$

Now ψ appears explicitly in (13) so p_ψ is not conserved. The conservation law we require is $p_\theta = \mu$, where

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = (m_1 + m_2) {r_1}^2 \dot{\theta} + m_2 {r_2}^2 (\dot{\theta} + \dot{\psi}) + m_2 r_1 r_2 (2\dot{\theta} + \dot{\psi}) \cos\psi. \tag{14}$$

From (14) we obtain the equation relating the differential forms,

$$\left([m_1 + m_2]{r_1}^2 + m_2 {r_2}^2 + 2 m_2 r_1 r_2 \cos\psi\right) d\theta + \left(m_2 r_1 r_2 \cos\psi + m_2 r_2^2\right) d\psi = \mu dt.$$

If we set $\mu = 0$ and let $m_1 = m_2 = m$, then we can immediately find the geometric phase for one full cycle of ψ:

$$\Delta\theta = -\int_0^{2\pi} \left(\frac{r_2^2 + r_1 r_2 \cos\psi}{2{r_1}^2 + {r_2}^2 + 2 r_1 r_2 \cos\psi}\right) d\psi. \tag{15}$$

This solution has some interesting features. First notice that in the special case of $r_1 = r_2 = r$, equation (15) becomes

$$\Delta\theta = -\int_0^{2\pi} \left(\frac{1 + \cos\psi}{3 + 2\cos\psi}\right) d\psi = -\left(1 - \frac{\sqrt{5}}{5}\right)\pi \approx -0.55\pi,$$

independently of r. This makes sense: if the outer arm spins all the way around relative to the inner arm, the inner arm must rotate in the other direction to conserve angular momentum. If the two arms coincide first when θ is zero, they will next coincide when $\theta \approx 100^\circ$. Or, if $r_1 = 2r_2$, the geometric phase over a full cycle of ψ is

$$\Delta\theta = -\int_0^{2\pi} \left(\frac{1 + 2\cos\psi}{9 + 4\cos\psi}\right) d\psi = -\left(1 - \frac{7}{\sqrt{65}}\right)\pi \approx -0.13\pi, \tag{16}$$

or about 24°. In general, for rational r_1/r_2, the change in θ for one closed loop in ψ is an irrational multiple of π.

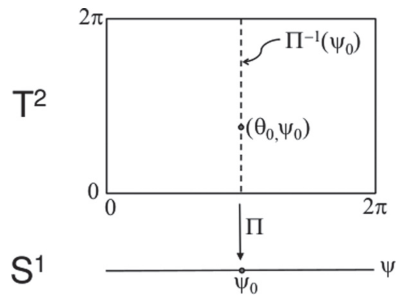


Figure 7 Schematic diagram of a reduction by symmetry $(\theta, \psi) \mapsto \psi$.

Reduction by symmetry and reconstruction

As we saw previously, in Elroy's beanie the configuration space M is the torus T^2 with local coordinates $\{\theta, \psi\}$, the symmetry group is $SO(2)$, whose action $SO(2) \times T^2 \to T^2$ is given by (9). The symmetry allowed us to simplify the problem, so we call that solution strategy ***reduction by symmetry***. Recovering the full solution is called ***reconstruction***. Let us illustrate how reduction and reconstruction work.

Reduction by symmetry. The energy of the system is the Lagrangian in (3) on the position-velocity space with coordinates $\{\theta, \psi, \dot{\theta}, \dot{\psi}\}$. Formally this space is called the ***tangent bundle*** *TM* of configuration space M, and it consists of the collection over all points $m \in M$ of the tangent space to M at m. Reduction by symmetry starts by identifying that the Lagrangian is invariant under the action in (9) lifted to *TM* to give

$$\alpha \cdot (\theta, \psi, \dot{\theta}, \dot{\psi}) = (\alpha + \theta, \psi, \dot{\theta}, \dot{\psi}).$$

(Observe that a change in the angle from which the laboratory frame views the boy-beanie system should have no effect upon the two bodies' angular velocities.) This action forms an equivalence relation on M,

$$m \sim m' \text{ if there exists an } \alpha \in G \text{ such that } m' = \alpha \cdot m.$$

We may define a natural projection to the quotient under $\sim$,

$$\Pi : M \to M/G = M/\sim,$$

which is sometimes called projection to the ***shape space*** M/G. There is a corresponding quotient

$$T\Pi : TM \to T(M/G).$$

In our case, the quotient map is given by

$$\Pi : T^2 \to T^2/SO(2) \simeq S^1 : (\theta, \psi) \mapsto \psi \tag{17}$$

with the tangent map

$$T\Pi : T(T^2) \to T(S^1) : (\theta, \psi, \dot{\theta}, \dot{\psi}) \mapsto (\psi, \dot{\psi}).$$

Because the energy is invariant under the action, an energy function can be defined unambiguously on the quotient. Indeed, the energy function "dropped" by the quotient map $TM \to T(M/G)$ is well defined. The reduced equations of motion are in the tangent bundle $T(M/G)$. In our case, they are simply $\ddot{\psi} = 0$, or uniform motion in the angle ψ.

Recall our choice of $p_\theta = \mu$ as the conservation law from which to derive the geometric phase. A more rigorous explanation of that choice comes from observing that p_θ is the momentum in the ***fiber***, the preimage of the projection map, $\Pi^{-1}(\psi_0) \simeq SO(2)$. The ***momentum mapping*** associated with the group action is the momentum induced by infinitesimally moving in the direction of the action, namely, the momentum p_θ in the fiber. The geometric phase now can be seen as a computation on a differential form defined on the zero-level set of the momentum mapping p_θ.

Reconstruction. Declaring $p_\theta = \mu = 0$ may seem dissatisfyingly ad hoc even though it makes physical sense. To further understand this choice, we need to learn a bit more about reconstruction.

Revisiting equation (11), we call the differential form on the right side the ***connection***

$$A = d\theta + \frac{I_2}{I_1 + I_2} d\psi \tag{18}$$

on T^2. Here's why. A differential form α is a dual tangent vector. That is, α acts on tangent vectors to T^2 (at a fixed point of T^2) by pairing. From the coordinates $\{\theta, \psi\}$ we construct a basis of the tangent space to T^2 at the point (θ_0, ψ_0),

$$\left\{ \partial_\theta := \left.\frac{\partial}{\partial \theta}\right|_{(\theta_0, \psi_0)}, \ \partial_\psi := \left.\frac{\partial}{\partial \psi}\right|_{(\theta_0, \psi_0)} \right\}.$$

Notice that $T\pi(\partial_\theta) = 0$ and $T\pi(\partial_\psi) = \left.\frac{\partial}{\partial \psi}\right|_{\psi_0}$. Inverting the projection map is not well defined, even if we insist that we stay on the fiber $\pi^{-1}(\psi_0)$. To stay on the fiber dictates that the vector field be of the form

$$\xi = \partial_\psi + f(\theta, \psi)\partial_\theta. \tag{19}$$

For a connection A, the ***horizontal lift*** is the unique vector field ξ on T^2 that satisfies both (19) and $A(\xi) = 0$. Then A has the form

$$A = d\theta - f(\theta, \psi) d\psi. \tag{20}$$

Thus, for the connection given in (18) the horizontal lift (19) is

$$\xi = \partial_\psi - \frac{I_2}{I_1 + I_2} \partial_\theta. \tag{21}$$

The reconstructed trajectories are the curves whose tangents are the horizontally lifted vector field. Notice that $f(\theta, \psi)$ in (19) is the instantaneous slope of the trajectory curve in M. In (21) the slope $f(\theta, \psi) = -I_2/(I_1 + I_2)$ is constant, and thus we reconfirm that the geometric phase is $\gamma = \Delta\theta = -2\pi I_2/(I_1 + I_2)$.

So how do we know that using this connection is the correct method to reconstruct the full solution in M? Consider *TM* to be a metric space with a metric at each point (θ, ψ) induced by the inner product

$$\langle \dot{\theta}, \dot{\psi} \rangle = L(\theta, \psi, \dot{\theta}, \dot{\psi}).$$

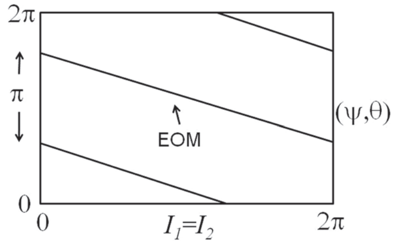


Figure 8 A reconstructed trajectory for Elroy's beanie with $I_1 = I_2$ and $\theta_0 = 5\pi/2$. The slope of the line is $-1/2$.

The trajectories are the integrals of the horizontally lifted vector fields which are precisely the geodesics of the kinetic energy metric—the curves that minimize the energy integral [**3**]. The ***connection*** is the device that calculates how to use the metric to perform the horizontal lift. To invert the projection of (θ, ψ) onto ψ (in Elroy notation), we simply choose an initial θ_0 when we construct the lifted curve. Notice the coefficient $f(\theta, \psi)$ in (19) is constant throughout T^2. Thus, the geodesics are straight lines in T^2. You can visualize each geodesic as a candy cane stripe running around the torus. Furthermore, if the ***curvature*** [**16**] of the connection (20) is

$$dA = d^2\theta - \frac{\partial f}{\partial \theta} d\theta \wedge d\psi = \frac{\partial f}{\partial \theta} d\psi \wedge d\theta,$$

where $\wedge$ denotes the exterior product of two differential forms, then the curvature of the connection (18) is zero—in other words, A is a ***flat*** connection. We can picture this because T^2 can be decomposed into a Cartesian product of the straight-line geodesics and the straight-line fibers of the projection. We can calculate this as well.

Returning to the teacup ride, the configuration space M is, again, the torus T^2 with local coordinates $\{\theta, \psi\}$. The free Lagrangian (13) is invariant under the action of $SO(2)$ in (9) lifted to TM, which allows the reduction in (17) and the accompanying tangent map. Again, p_θ in (14) is the momentum mapping, and the geometric phase will come from reconstruction using $p_\theta = 0$, leading us to a connection on T^2 for which the instantaneous slope function is now

$$f(\theta, \psi) = -\left(\frac{r_2^2 + r_1 r_2 \cos\psi}{2r_1^2 + r_2^2 + 2r_1 r_2 \cos\psi}\right),$$

which no longer is constant.

If $r_1 = 2r_2$, the differential form has slope function

$$f(\theta, \psi) = -\frac{1 + 2\cos\psi}{9 + 4\cos\psi}.$$

Thus the geometric phase is obtained correctly by (16). Now the geodesics of the kinetic energy metric are no longer straight lines in TM but the new connection on T^2 is still flat. (In general, $dA = 0$ if and only if $\frac{\partial f(\theta,\psi)}{\partial\theta} = 0$.) This illustrates that a flat space can be made of up an infinite collection of curved trajectories.

A falling cat

Cats are known to have an innate ability to change orientation as they fall so that they land on their feet. Let's consider the mechanics of the "righting-reflex" for a cat who

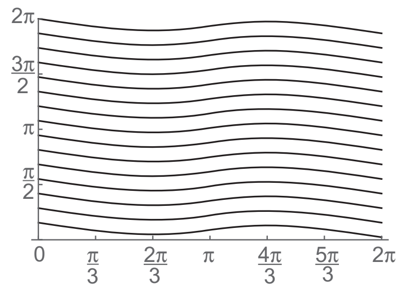


Figure 9 First 15 periods of a reconstructed trajectory for the teacup $r_1 = 2r_2$. The full trajectory fills the torus.

starts at rest upside down and twists her body to land on her feet. A simplified model of a falling cat [**12**] observes that she bends at the waist either to pull her head and tail closer together or to rotate her torso. By representing the body of the cat with a pair of identical right circular cylinders (representing the front and back of the cat) connected by a special "no-twist" joint which only allows these two motions, Montgomery [**18**] recast the problem in the language of geometric phase. Let ψ be the angle between the axes of rotation of the respective cylinders (so that $\psi/2$ is the angle of the front (or back) of the cat with respect to vertical). This is the angle of "back bend." The torso rotation can be modeled by allowing the cylinders to roll against each other. Let θ represent the angle of the roll, that is, if the front rolls by θ radians clockwise the back is constrained to roll by θ radians counterclockwise (see Figure 10).

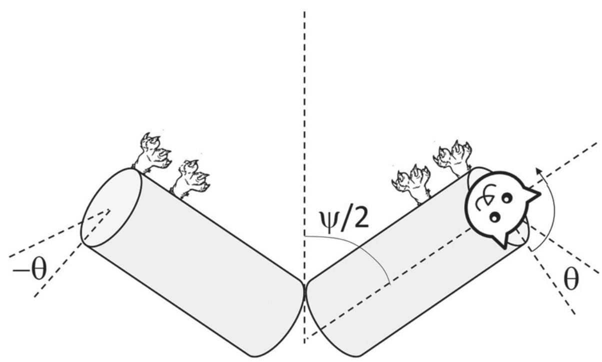


Figure 10 Falling cat and coordinate system.

The configuration space for the linked rigid cylinders can be identified as the group $SO(3)$ of rigid rotations in $\mathbb{R}^3$ with the coordinates $\{\theta, \psi, \chi\}$. In response to the cat changing shape by controlling θ and ψ, the entire body of the cat rotates by an angle χ (the angle that the plane of the two cylinder axes makes with positive vertical axis) presumably to land on her feet. Descriptively, the configuration space of the entire cat may be simplified by a reduction, using the symmetry in the variable χ. As a consequence, the two variables θ and ψ form the reduced space. (Technically, the reduced space is the real projective plane. See [**18**].) Conservation of angular momentum [**18**] thus determines a mechanical connection between the base variables θ and ψ which the cat can control, and the fiber variable χ, giving us a connection

$$A = d\chi - F(\theta, \psi, \chi)\, d\theta - G(\theta, \psi, \chi)\, d\psi.$$

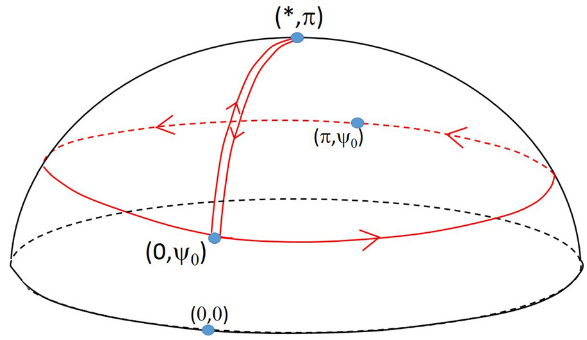


Figure 11 The correspondence between the falling cat and parallel transport. The polar angle is θ, and the azimuthal angle is $\phi = \pi/2 - \psi/2$.

Note that a rotation in the angle ψ happens in the plane of the cat and has no effect on the angular momentum about the horizontal axis that passes through the plane of the cat. Thus $G = 0$. Through calculating moments of inertia, it can be shown [**18**] that $F = \sin(\psi/2)$. Thus the connection is

$$A = d\chi - \sin(\psi/2)\, d\theta,$$

and the geometric phase may be computed by

$$\Delta\chi = \int_C \sin(\psi/2)\, d\theta.$$

Consider a basic trajectory for the cat [**21**], in which the cat starts with her feet up and back straight. She then bends her back to move the front and back cylinders toward each other to get to a fixed value $\psi = \psi_0$. Holding that angle constant, she now performs a full rotation in θ, and then straightens back out again. Since we desire the cat to land on her feet, we seek a geometric phase of $\Delta\chi = \pi$ radians. This gives us

$$\begin{aligned} \Delta\chi &= \int_0^{2\pi} \sin(\psi_0/2)\, d\theta \\ \pi &= 2\pi \sin(\psi_0/2) \\ \frac{\psi_0}{2} &= \sin^{-1}\left(\frac{1}{2}\right) \\ \psi_0 &= \frac{\pi}{3}. \end{aligned}$$

So the cat needs to bend her back 60°, do a full twist, and then straighten back out to land upright. This also may determine whether or not the cat succeeds in landing on her feet. If she does not have enough time to execute a full rotation in θ, then she under-rotates since $\Delta\chi < \pi$.

Note the correspondence between the falling cat and parallel transport on the sphere. In standard spherical coordinates, the polar angle is θ, and the azimuthal angle is $\phi = \pi/2 - \psi/2$, so that a stretched cat with $\psi = \pi$ and θ undefined starts at the North Pole. The one form

$$F(\theta, \psi, \chi)\, d\theta = \sin(\psi/2)\, d\theta = \sin(\pi/2 - \phi)\, d\theta = \cos\phi\, d\theta$$

has curvature

$$d(F(\theta, \psi, \chi)d\theta) = -\sin\phi\, d\phi \wedge d\theta \;= \sin\phi\, d\theta \wedge d\phi,$$

which is exactly the differential element of area in spherical coordinates. This is also the measure of precession of a planar Foucault pendulum as it takes this path around the sphere [**4, 20**].

There is a connection (pun intended!) between the falling cat and Elroy's beanie. Suppose that Elroy spends a lot of time spacewalking. In order to turn he spins his propeller, causing his body to rotate in the opposite direction. When he achieves the desired angle, he reaches up and stops the propeller, which stops his spin. Elroy and his beanie, taken together, constitute a nonrigid body that is very similar to the falling cat: Elroy controls the difference angle ψ, just as the cat controls the two angles θ and ψ. As a consequence, Elroy's orientation θ has changed, just as the cat's χ angle has changed. If Elroy were horizontal and falling, he could right himself by controlling ψ. That is, he could go from face-up to face-down as long as the geometric phase between ψ and θ was π or $-\pi$.

Conclusion

We have seen how geometric phase arises in several contexts, and we have shown that in some cases we may compute the geometric phase without necessarily knowing the equations of motion of the system. As introduced before, Berry's idea of phase was revolutionary: it enabled many problems in quantum mechanics to be solved more simply, on a descriptive level. For example, a complete explanation of electric polarization in dielectrics (electrical insulators) remained elusive until Berry's phase was incorporated [**13**]. But geometric phase is also an effective tool in classical mechanical problems that we can observe every day.

REFERENCES

1. Y. Aharonov, D. Bohm, Significance of electromagnetic potentials in quantum theory, *Phys. Rev.* **115** (1959) 485–491, http://dx.doi.org/10.1103/PhysRev.115.485.
2. J. Anandan, J. Christian, K. Wanelik, Resource Letter GPP-1: Geometric phases in physics, *Am. J. Phys.* **65** no. 3 (1997) 180–185, http://dx.doi.org/10.1119/1.18570.
3. V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*. Corrected reprint of the second edition. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein. Graduate Texts in Mathematics, Vol. 60, Springer, New York, 1989, http://dx.doi.org/10.1007/978-1-4757-2063-1.
4. J. von Bergmann, H. von Bergmann, Foucault pendulum through basic geometry, *Am. J. Phys.* **75** no. 10 (2007) 888–892, http://dx.doi.org/10.1119/1.2757623.
5. M. Berry, Quantal phase factors accompanying adiabatic changes, *Proc. R. Soc. London Ser. A Math. Phys. Eng. Sci.* **392** (1984) 45–57, http://dx.doi.org/10.1098/rspa.1984.0023.
6. M. Berry, The geometric phase, *Sci. Amer.* **259** no. 6 (1988) 26–34.
7. A. Bloch, *Nonholonomic Mechanics and Control*. Springer, New York, 2003, http://dx.doi.org/10.1007/b97376.
8. D. Chruściński, A. Jamiołkowski, *Geometric Phases in Classical and Quantum Mechanics*. Progress in Mathematical Physics, Vol. 36, Birkhäuser, Boston, 2004, http://dx.doi.org/10.1007/978-0-8176-8176-0.
9. J. Hannay, Angle variable holonomy in adiabatic excursion of an integrable Hamiltonian, *J. Phys. A: Math. Gen.* **18** (1985) 221–230, http://dx.doi.org/10.1088/0305-4470/18/2/011.
10. J. Hart, R. Miller, K. Wanelik, A simple geometric model for visualizing the motion of a Foucault pendulum, *Am. J. Phys.* **55** no. 1 (1987) 67–70, http://dx.doi.org/10.1119/1.14972.
11. D. Holm, *Geometric Mechanics, Part I: Dynamics and Symmetry*. Imperial College Press, London, 2008, http://dx.doi.org/10.1142.
12. T. R. Kane, M. P. Scher, A dynamical explanation of the falling cat phenomenon, *Int. J. Solids Struct.* **55** (1969) 663–670, http://dx.doi.org/10.1016/0020-7683(69)90086-9.

13. R. King-Smith, D. Vanderbilt, Theory of polarization of crystalline solids, *Phys. Rev. B* **47** (1993) 1651–1654, http://dx.doi.org/10.1103/PhysRevB.47.1651.
14. D. Little, SpiroGraph (2006), http://www.math.psu.edu/dlittle/java/parametricequations/spirograph/.
15. J. Marsden, R. Montgomery, T. Ratiu, Reduction, symmetry, and phases in mechanics, *Mem. Amer. Math. Soc.* **88** (1990), http://dx.doi.org/10.1090/memo/0436.
16. J. Marsden, *Lectures on Mechanics*. London Mathematical Society Lecture Note Series, Vol. 174, Cambridge Univ. Press, Cambridge, 1992, http://dx.doi.org/10.1017/CBO9780511624001.
17. J. Marsden, T. Ratiu, *Introduction to Mechanics and Symmetry*. Second edition. Springer-Verlag, New York, 1999, http://dx.doi.org/10.1017/CBO9780511624001.
18. R. Montgomery, Gauge theory of the falling cat, in *Dynamics and Control of Mechanical Systems*. Ed. M. J. Enos. American Mathematical Society, Providence, RI, 1993. 193–218.
19. A. Mostafazadeh, *Dynamical Invariants, Adiabatic Approximation and the Geometric Phase*. Nova Science, New York, 2002.
20. J. Oprea, Geometry and the Foucault pendulum, *Amer. Math. Monthly* **102** no. 6 (1995) 515–522, http://dx.doi.org/10.2307/2974765.
21. G. J. Rademaker, J. W. G. ter Braak, Das Umdrehen der fallenden Katze in der Luft, *Acta oto-lar.* **23** (1935) 313–343.
22. R. Resta, Manifestations of Berry's phase in molecules and condensed matter, *J. Phys.: Condens. Matter* **12** no. 9 (2000) R107–R143, http://dx.doi.org/10.1088/0953-8984/12/9/201.
23. A. Shapere, F. Wilczek, *Geometric Phases in Physics*. World Scientific, Singapore, 1989, http://dx.doi.org/10.1142/9789812798381.
24. E. Weisstein, Berry Phase—From ScienceWorld, A Wolfram Web Resource, http://scienceworld.wolfram.com/physics/BerryPhase.html.

Summary. Geometric phase in a dynamical system can be visualized as the interplay between two periodic functions which go in and out of "synch." Using illustrations of a boy's walk in space and a dizzying fun park ride, we demonstrate that in certain simple mechanical systems we can compute geometric phase directly from a symmetry—we don't even need to solve the system of differential equations. We can also use geometric phase to explain how cats (almost) always land on their feet. We conclude with an interpretation of geometric phase in terms of the geometric notions of connection and curvature.

JEFFREY LAWSON (MR Author ID: 352605) is a professor and the head of the Department of Mathematics and Computer Science at Western Carolina University. He has also served on the faculty at Trinity University and Wake Forest University. He earned a Ph.D. in mathematics from North Carolina State University. His research is in geometric mechanics and multisymplectic geometry. He enjoys mountain biking with his family, hiking, running, and reading. Contrary to what this article may imply, he does not drop his cats from large heights.

MATTHEW RAVE (MR Author ID: 860526) earned a Ph.D. in physics from Wake Forest University in 2007 before moving to the mountains of North Carolina, where he is an associate professor of physics at Western Carolina University. His research interests (besides the geometric phase) include 1D and 2D solid state systems, foundational questions about quantum interference and decoherence, and physics education. He enjoys reading, writing (he maintains a physics blog), cooking, and (according to his wife Sarah) playing way too much chess.

Bounds for the Representations of Integers by Positive Quadratic Forms

KENNETH S. WILLIAMS
Carleton University
Ottawa, ON, Canada
kennethwilliams@cunet.carleton.ca

The fact that every positive integer is the sum of four squares of integers was first proved in 1770 by Lagrange [**9**]. For example the integer 1770 is $1^2 + 1^2 + 2^2 + 42^2$. We can view Lagrange's theorem as telling us that the polynomial $x_1^2 + x_2^2 + x_3^2 + x_4^2$ in the four real variables x_1, x_2, x_3 and x_4 represents every positive integer since whatever positive integer n is specified there are integer values of x_1, x_2, x_3 and x_4, say y_1, y_2, y_3, and y_4, respectively, such that $y_1^2 + y_2^2 + y_3^2 + y_4^2 = n$. The quadruple (y_1, y_2, y_3, y_4) is called the representation of n by the polynomial $x_1^2 + x_2^2 + x_3^2 + x_4^2$. The polynomial $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is an example of a quadratic form as it has the property that if we replace each of x_1, x_2, x_3, x_4 by tx_1, tx_2, tx_3, tx_4, respectively, where t is a real number, we obtain t^2 times the original polynomial. (All quadratic forms in this article are assumed to have integer coefficients.) Moreover, the quadratic form $x_1^2 + x_2^2 + x_3^2 + x_4^2$ has the additional property that it is nonnegative for all real values of x_1, x_2, x_3, and x_4 and is zero only when x_1, x_2, x_3, and x_4 are all zero. Quadratic forms with this property are called positive. (Frobenius gave in 1894 necessary and sufficient conditions for a quadratic form to be positive, see for example [**11**, Theorem 13.3.1, p. 400].) A positive quadratic form is called universal if it represents every positive integer. Thus $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is an example of a universal positive quadratic form.

The problem of determining whether a given positive quadratic form is universal has been solved in some ground-breaking work in recent years. It was first solved by Conway and Schneeberger [**5**] for positive quadratic forms all of whose cross-product terms have even coefficients. (For example the cross-product terms of the positive quadratic form $3x_1^2 + 3x_2^2 + 3x_3^2 + 2x_1x_2 + 2x_1x_3$ are all even, namely 2, 2, and 0.) They proved but did not publish the proof of the following theorem.

15-Theorem. *Let f be a positive quadratic form in any number of variables such that all the coefficients of its cross-product terms are even integers. If f represents all the positive integers up to and including* 15 *then f is universal.*

The details of their work are given in [**5**] and [**13**]. In 2000 Bhargava [**2**] gave a beautiful new proof of this theorem in the following stronger form.

Strong 15-Theorem. *Let f be a positive quadratic form in any number of variables such that all the coefficients of its cross-product terms are even integers. If f represents all the nine integers*

$$1, 2, 3, 5, 6, 7, 10, 14, 15$$

then it is universal.

Bhargava also showed that the set $\{1, 2, 3, 5, 6, 7, 10, 14, 15\}$ is minimal in a certain sense.

Math. Mag. **89** (2016) 122–131. doi:10.4169/math.mag.89.2.122.
MSC: Primary 11E20, Secondary 15A15

In 1993 Conway formulated the conjecture that a positive quadratic form that represents all the positive integers up to and including 290 must be universal. This was proved by Bhargava and Hanke [**3**].

290-Theorem. *If a positive quadratic form in any number of variables represents all the positive integers up to and including* 290 *then it is universal.*

Indeed Bhargava and Hanke proved this result in the following stronger form.

Strong 290-Theorem. *If a positive quadratic form in any number of variables represents all the twenty-nine integers*

$$1, 2, 3, 5, 6, 7, 10, 13, 14, 15, 17, 19, 21, 22, 23, 26, 29,$$

$$30, 31, 34, 35, 37, 42, 58, 93, 110, 145, 203, 290$$

then it is universal.

Bhargava and Hanke showed too that the set in the strong 290-theorem is also minimal.

The beauty of these results is that given a positive quadratic form we have only to check that it represents nine integers to know that it represents all positive integers in the "all cross-product coefficients even" case and twenty-nine integers in the case that the form has at least one odd cross-product coefficient. This leads us naturally to the question "How do we check if a positive quadratic form represents a certain positive integer?" The main result of this article provides a simple answer to this question. Let us consider an example for guidance. We show for the quadratic form considered in this example that for each positive integer n there are only finitely many representations of n by the form and moreover that all such representations lie in a certain hypercube.

Example. We consider the quadratic form

$$\begin{aligned} G(x_1, x_2, x_3, x_4) := 3x_1^2 + 44x_2^2 + 13x_3^2 + 18x_4^2 + 2x_1x_2 + 6x_1x_3 + 8x_1x_4 \\ + 42x_2x_3 + 16x_2x_4 + 8x_3x_4. \end{aligned} \tag{1}$$

We note that in this example all the cross-product terms of G have even coefficients. We leave it to the reader to check that G satisfies the determinantal conditions of Frobenius so that G is a positive form. Alternatively we can see that G is a positive form from the identity

$$\begin{aligned} 4323G = 1441(3x_1 + x_2 + 3x_3 + 4x_4)^2 + 11(131x_2 + 60x_3 + 20x_4)^2 \\ + 30(11x_3 - 40x_4)^2 + 2358x_4^2, \end{aligned}$$

which was found by first completing the square in x_1 in G, then the square in x_2 and finally the square in x_3. (The coefficients of the squares in this identity are all positive and G can only be zero when

$$3x_1 + x_2 + 3x_3 + 4x_4 = 131x_2 + 60x_3 + 20x_4 = 11x_3 - 40x_4 = x_4 = 0,$$

that is when $x_1 = x_2 = x_3 = x_4 = 0$, so G is positive.)

Our idea is to attempt to express G in the form

$$\begin{aligned}G(x_1,x_2,x_3,x_4) &= G(0, x_2 - tx_1, x_3 - ux_1, x_4 - vx_1) + wx_1^2\\ &= 44(x_2 - tx_1)^2 + 13(x_3 - ux_1)^2 + 18(x_4 - vx_1)^2\\ &\quad + 42(x_2 - tx_1)(x_3 - ux_1) + 16(x_2 - tx_1)(x_4 - vx_1)\\ &\quad + 8(x_3 - ux_1)(x_4 - vx_1) + wx_1^2\end{aligned}$$

for some rational numbers t, u, v, and w. Clearly the coefficients of x_2^2, x_3^2, x_4^2, x_2x_3, x_2x_4, and x_3x_4 agree so we have only to arrange the agreement of the coefficients of x_1^2, x_1x_2, x_1x_3, and x_1x_4. Equating the coefficients of x_1x_2, x_1x_3, and x_1x_4, we obtain

$$\begin{aligned}2 &= -88t - 42u - 16v,\\ 6 &= -42t - 26u - 8v,\\ 8 &= -16t - 8u - 36v.\end{aligned}$$

Solving these linear equations for t, u, and v we obtain

$$t = \frac{150}{361}, \quad u = -\frac{301}{361}, \quad v = -\frac{80}{361}.$$

Equating the coefficients of x_1^2, we have

$$3 = 44t^2 + 13u^2 + 18v^2 + 42tu + 16tv + 8uv + w.$$

Using the values of t, u, and v in this equation, we deduce $w = \frac{10}{361}$. This shows that

$$\begin{aligned}G = {} & 44\left(x_2 - \frac{150}{361}x_1\right)^2 + 13\left(x_3 + \frac{301}{361}x_1\right)^2 + 18\left(x_4 + \frac{80}{361}x_1\right)^2\\ & + 42\left(x_2 - \frac{150}{361}x_1\right)\left(x_3 + \frac{301}{361}x_1\right) + 16\left(x_2 - \frac{150}{361}x_1\right)\left(x_4 + \frac{80}{361}x_1\right)\\ & + 8\left(x_3 + \frac{301}{361}x_1\right)\left(x_4 + \frac{80}{361}x_1\right) + \frac{10}{361}x_1^2.\end{aligned}$$

Thus, if n is a positive integer which is represented by G, then there are integers y_1, y_2, y_3, and y_4 such that $G(y_1, y_2, y_3, y_4) = n$ and our expression for G gives

$$n = G(0, y_2 - \frac{150}{361}y_1, y_3 + \frac{301}{361}y_1, y_4 + \frac{80}{361}y_1) + \frac{10}{361}y_1^2.$$

As $G(x_1, x_2, x_3, x_4)$ is a positive quadratic form so is $G(0, x_2, x_3, x_4)$ and we have

$$G(0, y_2 - \frac{150}{361}y_1, y_3 + \frac{301}{361}y_1, y_4 + \frac{80}{361}y_1) \geq 0$$

and thus

$$n \geq \frac{10}{361}y_1^2.$$

This shows that for a fixed positive integer n there are only finitely many possibilities for the integer y_1 given by

$$|y_1| \leq \sqrt{\frac{361}{10}n}.$$

Similarly we obtain

$$\begin{aligned} G(x_1, x_2, x_3, x_4) &= G(x_1 - \frac{45}{19}x_2, 0, x_3 + 2x_2, x_4 + \frac{10}{19}x_2) + \frac{3}{19}x_2^2, \\ &= G(x_1 + \frac{129}{109}x_3, x_2 + \frac{380}{763}x_3, 0, x_4 - \frac{200}{763}x_2) + \frac{30}{763}x_3^2, \\ &= G(x_1 + \frac{48}{11}x_4, x_2 + \frac{20}{11}x_4, x_3 - \frac{40}{11}x_4) + \frac{6}{11}x_4^2, \end{aligned}$$

so that

$$|y_2| \leq \sqrt{\frac{19}{3}n}, \quad |y_3| \leq \sqrt{\frac{763}{30}n}, \quad |y_4| \leq \sqrt{\frac{11}{6}n}.$$

In this example we have shown that there are only finitely many solutions in integers y_1, y_2, y_3, y_4 to $G(y_1, y_2, y_3, y_4) = n$ for a given positive integer n, and every such solution vector (y_1, y_2, y_3, y_4) lies in the hypercube

$$|y_1| \leq \sqrt{\frac{361}{10}n}, \quad |y_2| \leq \sqrt{\frac{19}{3}n}, \quad |y_3| \leq \sqrt{\frac{763}{30}n}, \quad |y_4| \leq \sqrt{\frac{11}{6}n}.$$

The main result of this article is the explicit determination of the corresponding hypercube for a general positive quadratic form. Thus, in view of the 15- and 290-Theorems, we have only to check the finite number of values of a given positive quadratic form on the integral points of a hypercube to determine its universality or nonuniversality. Before proving this result we make a few remarks about universal and nonuniversal positive quadratic forms.

Universal and nonuniversal forms

Clearly a positive quadratic form in one variable cannot be universal. Such a form is $a_1x_1^2$, where a_1 is a positive integer, and $a_1x_1^2$ cannot represent the positive integer $2a_1$. What about positive quadratic forms in two variables? Such forms in two variables are called binary. The general binary quadratic form is $a_1x_1^2 + a_{12}x_1x_2 + a_2x_2^2$, where a_1, a_{12} and a_2 are integers. As

$$a_1x_1^2 + a_{12}x_1x_2 + a_2x_2^2 = \frac{1}{4a_1}(2a_1x_1 + a_{12}x_2)^2 + \frac{(4a_1a_2 - a_{12}^2)}{4a_1}x_2^2,$$

we see that the form is positive if and only if

$$a_1 > 0, \quad 4a_1a_2 - a_{12}^2 > 0.$$

Proposition. *A positive binary quadratic form $a_1x_1^2 + a_{12}x_1x_2 + a_2x_2^2$ is never universal.*

Proof. The integer $d := a_{12}^2 - 4a_1a_2$ is strictly negative so it is not a perfect square. Thus, from the theory of quadratic residues modulo a prime, we know that there are infinitely many primes p such that d is a quadratic nonresidue modulo p. Hence we can choose the prime p to satisfy $p > 4a_1|d|$. Suppose that $a_1x_1^2 + a_{12}x_1x_2 + a_2x_2^2$ is universal. Then it represents p and so there are integers u and v such that

$$p = a_1u^2 + a_{12}uv + a_2v^2.$$

Hence

$$4a_1 p = x^2 - dy^2, \text{ where } x := 2a_1 u + a_{12} v \text{ and } y := v.$$

If p divides y then, from $4a_1 p = x^2 - dy^2$, we see that p divides x. Thus p^2 divides $x^2 - dy^2 = 4a_1 p$. This is impossible as $p > 4a_1$. Hence p does not divide y. Thus there is an integer z such that $yz \equiv 1 \pmod{p}$. Then $(xz)^2 \equiv d \pmod{p}$, contradicting that d is a quadratic nonresidue modulo p. Therefore, $a_1 x_1^2 + a_{12} x_1 x_2 + a_2 x_2^2$ cannot be universal. ■

What about a positive quadratic form in three variables? Such forms are called ternary quadratic forms and they like binary quadratic forms can never be universal. To prove that a general positive ternary quadratic form is never universal is much more difficult than in the binary case. Proofs can be found in Albert [**1**, p. 291, Theorem 13] and Conway [**4**, p. 142]. In the diagonal case, which is when there are no cross-product terms in the ternary form (so that the ternary form is $a_1 x_1^2 + a_2 x^2 + a_3 x_3^2$), a simple proof has been given by Dickson [**8**, p. 104, Theorem 95].

What about positive quadratic forms in four variables? These forms are called quaternary. Can they represent all positive integers? Since $a^2 + b^2 \equiv 0, 1, 2 \pmod{4}$ for any integers a and b, the form $x_1^2 + x_2^2 + 4a_3 x_3^2 + 4a_4 x_4^2$, where a_3 and a_4 are positive integers, cannot represent any positive integer $n \equiv 3 \pmod{4}$. Thus there are infinitely many positive quaternary quadratic forms which are not universal. On the other hand, Liouville and other mathematicians showed that there are positive quaternary quadratic forms different from $x_1^2 + x_2^2 + x_3^2 + x_4^2$, which are universal. An example of Liouville [**10**, p. 271] follows easily from Lagrange's theorem. Let n be a positive integer and let a, b, c, and d be integers such that $n = a^2 + b^2 + c^2 + d^2$. Since there are only two possible residues for an integer modulo 2, namely 0 and 1, by Dirichlet's box principle at least two of b, c, and d must have the same residue modulo 2, say $c \equiv d \pmod{2}$. Then $n = a^2 + b^2 + 2e^2 + 2f^2$, where e and f are the integers $(c + d)/2$ and $(c - d)/2$, respectively. Thus the quadratic form $x_1^2 + x_2^2 + 2x_3^2 + 2x_4^2$ is universal. Ramanujan [**12**] and Dickson [**6, 8**] determined all the universal positive quaternary quadratic forms $a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2 + a_4 x_4^2$, where a_1, a_2, a_3, and a_4 are positive integers satisfying $a_1 \le a_2 \le a_3 \le a_4$. They proved that there are precisely 54 such forms. The interested reader can find them listed in Dickson [**8**, p. 105]. Many examples of universal positive quaternary quadratic forms with nonzero cross-product terms have been given by Dickson, see Dickson [**7**].

For forms in more than four variables there are infinitely many such forms which are universal, for example $x_1^2 + x_2^2 + x_3^2 + x_4^2 + a_5 x_5^2 + \cdots + a_n x_n^2$, where $a_5, \ldots, a_n$ are positive integers, and infinitely many forms which are not, for example $x_1^2 + x_2^2 + 4a_3 x_3^2 + \cdots + 4a_n x_n^2$, where $a_3, \ldots, a_n$ are positive integers.

These examples show the wide variety of possibilities for the universality or nonuniversality of positive quadratic forms and so demonstrate the simplicity and power of the 15- and 290-Theorems.

Notation and Main Result

We now state and prove the central result of this article, which we have not found in the literature. We denote the sets of integers, positive integers and rational numbers by $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{Q}$, respectively. The general quadratic form with integer coefficients in k real variables $x_1, \ldots, x_k$ is given by

$$F(x_1, \ldots, x_k) := \sum_{1 \le i \le j \le k} a_{ij} x_i x_j, \quad a_{ij} \in \mathbb{Z}. \tag{2}$$

We assume that F is positive so that the coefficients a_{ij} satisfy the previously mentioned conditions of Frobenius. Let n be a positive integer. Our objective is to give an explicit hypercube in $\mathbb{Z}^k$ in which all the solutions in integers of $F(x_1, \ldots, x_k) = n$ (if any) must lie.

The matrix of the form F is the $k \times k$ symmetric matrix

$$A := \begin{bmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1k} \\ \frac{1}{2}a_{12} & a_{22} & \cdots & \frac{1}{2}a_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \frac{1}{2}a_{1k} & \frac{1}{2}a_{2k} & \cdots & a_{kk} \end{bmatrix},$$

where each entry of A is either an integer or half an odd integer. Thus for the quadratic form G in the example (see (1)) we have

$$A = \begin{bmatrix} 3 & 1 & 3 & 4 \\ 1 & 44 & 21 & 8 \\ 3 & 21 & 13 & 4 \\ 4 & 8 & 4 & 18 \end{bmatrix}.$$

If all the coefficients of the cross-product terms in F are even, then all the entries in A are integers and the form F is said to be an integer-matrix form. The form F and its matrix A are related by $F = X^t AX$, where ^{t} denotes the transpose of a matrix and $X = [x_1 \cdots x_k]^t$. As F is a positive quadratic form, and A is the matrix of F, by the determinantal conditions of Frobenius we have $\det A > 0$. In the case of G we have $\det A = 60$.

For $j = 1, 2, \ldots, k$ with $k \geq 2$ we let A_j denote the $(k-1) \times (k-1)$ symmetric matrix formed by deleting the jth row and jth column of A. The principal diagonal of A_j $(j = 1, 2, \ldots, k)$ is part of the principal diagonal of A so, as F is a positive quadratic form with matrix A, again by the Frobenius determinantal conditions for a quadratic form to be positive, we have $\det A_j > 0$ $(j = 1, 2, \ldots, k)$. In the case of G we have

$$A_1 = \begin{bmatrix} 44 & 21 & 8 \\ 21 & 13 & 4 \\ 8 & 4 & 18 \end{bmatrix}, A_2 = \begin{bmatrix} 3 & 3 & 4 \\ 3 & 13 & 4 \\ 4 & 4 & 18 \end{bmatrix}, A_3 = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 44 & 8 \\ 4 & 8 & 18 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} 3 & 1 & 3 \\ 1 & 44 & 21 \\ 3 & 21 & 13 \end{bmatrix},$$

and

$$\det A_1 = 2166,\ \det A_2 = 380,\ \det A_3 = 1526,\ \det A_4 = 110.$$

We prove our main result by using some elementary computational matrix algebra to express $F(x_1, \ldots, x_k)$ for $r = 1, \ldots, k$ in the form

$$F(x_1 - t_1x_r, \ldots, x_{r-1} - t_{r-1}x_r, 0, x_{r+1} - t_{r+1}x_r, \ldots, x_k - t_kx_r) + u_rx_r^2$$

for some rational numbers $t_1, \ldots, t_{r-1}, t_{r+1}, \ldots, t_k$ and an explicitly known rational number u_r as suggested by the example.

Theorem. *Let $k \in \mathbb{N}$ be such that $k \geq 2$. Let F be the integral positive quadratic form in the k indeterminates $x_1, \ldots, x_k$ given in (2). Let $n \in \mathbb{N}$. Then there are only finitely* *many solutions* $(y_1, \ldots, y_k) \in \mathbb{Z}^k$ *of*

$$F(y_1, \ldots, y_k) = n$$

and any such solution satisfies

$$|y_r| \le \left\lfloor \sqrt{\frac{\det A_r}{\det A}}\sqrt{n} \right\rfloor, \quad r = 1, 2, \ldots, k,$$

where $\lfloor x \rfloor$ *denotes the floor of the real number* x.

Proof. Fix $r \in \{1, 2, \ldots, k\}$. As $\det A_r \neq 0$ the system of $k-1$ linear equations in the $k-1$ unknowns $t_1, \ldots, t_{r-1}, t_{r+1}, \ldots, t_k$

$$A_r \begin{bmatrix} t_1 \\ \vdots \\ t_{r-1} \\ t_{r+1} \\ \vdots \\ t_k \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}a_{1\,r} \\ \vdots \\ -\frac{1}{2}a_{r-1\,r} \\ -\frac{1}{2}a_{r\,r+1} \\ \vdots \\ -\frac{1}{2}a_{r\,k} \end{bmatrix} \tag{3}$$

has a unique solution $(t_1, \ldots, t_{r-1}, t_{r+1}, \ldots, t_k) \in \mathbb{Q}^{k-1}$. Let T be the $k \times k$ matrix defined by

$$T := \begin{bmatrix} 1 & 0 & \cdots & 0 & t_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & t_{r-1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & t_{r+1} & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \ldots & 0 & t_k & 0 & \cdots & 1 \end{bmatrix},$$

where the column with first entry t_1 is the rth column. Expanding the determinant of T by the rth row we obtain $\det T = 1$.

Suppose now that $(x_1, \ldots, x_k) \in \mathbb{Z}^k$ is a solution of $F(x_1, \ldots, x_k) = n$. For $i = 1, 2, \ldots, k$ define $y_i \in \mathbb{Q}$ by

$$y_i := \begin{cases} x_i - t_i x_r & \text{if } i \neq r, \\ x_r & \text{if } i = r. \end{cases} \tag{4}$$

Let

$$X := \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, \quad Y := \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix},$$

so that $n = X^t A X$ and $X = TY$. Hence

$$n = (TY)^t A (TY) = Y^t (T^t A T) Y. \tag{5}$$

We now determine the matrix T^tAT. The matrix T^tA is just the matrix A with the rth row replaced by $[u_1\, u_2 \cdots u_k]$, where for $i = 1, 2, \dots, k$

$$u_i = \begin{bmatrix} t_1 & \cdots & t_{r-1} & 1 & t_{r+1} & \cdots & t_k \end{bmatrix} \begin{bmatrix} \frac{1}{2}a_{1\,i} & \cdots & \frac{1}{2}a_{i-1\,i} & a_{i\,i} & \frac{1}{2}a_{i\,i+1} & \cdots & \frac{1}{2}a_{i\,k} \end{bmatrix}^t .$$

If $i = r$ we have

$$u_r = \sum_{j=1}^{r-1} \frac{1}{2} a_{j\,r} t_j + a_{r\,r} + \sum_{j=r+1}^{k} \frac{1}{2} a_{r\,j} t_j .$$

If $i < r$ we have

$$u_i = \sum_{j=1}^{i-1} \frac{1}{2} a_{j\,i} t_j + a_{i\,i} t_i + \sum_{j=i+1}^{r-1} \frac{1}{2} a_{i\,j} t_j + \frac{1}{2} a_{i\,r} + \sum_{j=r+1}^{k} \frac{1}{2} a_{i\,j} t_j = 0,$$

appealing to (3). Similarly if $i > r$ we have

$$u_i = \sum_{j=1}^{r-1} \frac{1}{2} a_{j\,i} t_j + \frac{1}{2} a_{r\,i} + \sum_{j=r+1}^{i-1} \frac{1}{2} a_{j\,i} t_j + a_{i\,i} t_i + \sum_{j=i+1}^{k} \frac{1}{2} a_{i\,j} t_j = 0,$$

again appealing to (3). Thus the rth row of T^tA is $\begin{bmatrix} 0 & \cdots & 0 & u_r & 0 & \cdots & 0 \end{bmatrix}$. Hence the matrix T^tAT is the same as the matrix A except that the rth row is $[0 \cdots 0\ u_r\ 0 \cdots 0]$ and the rth column is $[0 \cdots 0\ u_r\ 0 \cdots 0]^t$. Thus by (5) we have

$$\begin{aligned} n = Y^t(T^tAT)Y &= \sum_{\substack{1 \le i \le j \le k \\ i,\, j \ne r}} a_{i\,j} y_i y_j + u_r y_r^2 \\ &= F(y_1, \dots, y_{r-1}, 0, y_{r+1}, \dots, y_k) + u_r y_r^2 . \end{aligned} \tag{6}$$

As F is a positive form, we have $F(y_1, \dots, y_{r-1}, 0, y_{r+1}, \dots, y_k) \ge 0$, and thus from (4) and (6), we deduce

$$n \ge u_r x_r^2 . \tag{7}$$

All that remains is to determine u_r. Expanding $\det(T^tAT)$ by the rth row, we deduce that

$$\det(T^tAT) = u_r \det A_r .$$

As $\det T = 1$ we have

$$\det(T^tAT) = \det(T^t) \det A \det T = \det A .$$

Hence

$$u_r = \frac{\det A}{\det A_r} . \tag{8}$$

Finally, from (7) and (8), we obtain

$$|x_r| \le \sqrt{\frac{\det A_r}{\det A}} \sqrt{n}, \quad r = 1, 2, \dots, k,$$

from which the theorem follows. ■

We remark that the theorem is best possible in the sense that the hypercube cannot in general be made smaller and still contain all the solutions $(y_1, \ldots, y_k) \in \mathbb{Z}^k$ of $F(y_1, \ldots, y_k) = n$. To see this take for example $F(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ (so that $k = 3$) and $n = 3$. Here

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_1 = A_2 = A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

so that $\det A = \det A_1 = \det A_2 = \det A_3 = 1$ and as $\lfloor \sqrt{3} \rfloor = 1$ the hypercube is

$$|y_r| \leq 1, \quad r = 1, 2, 3.$$

The only solutions of $y_1^2 + y_2^2 + y_3^2 = 3$ are $(y_1, y_2, y_3) = (\pm 1, \pm 1, \pm 1)$ (8 choices of sign) and all of these lie on the boundary of the hypercube.

Example (continued). We apply our theorem in conjunction with the strong 15-Theorem to show that G (defined in (1)) is universal. By the theorem any solution $(y_1, y_2, y_3, y_4) \in \mathbb{Z}^4$ of $G(y_1, y_2, y_3, y_4) = n$ must satisfy

$$|y_1| \leq \left\lfloor \sqrt{\frac{\det A_1}{\det A}} \sqrt{n} \right\rfloor = \left\lfloor \sqrt{\frac{361}{10}} \sqrt{n} \right\rfloor,$$

$$|y_2| \leq \left\lfloor \sqrt{\frac{\det A_2}{\det A}} \sqrt{n} \right\rfloor = \left\lfloor \sqrt{\frac{19}{3}} \sqrt{n} \right\rfloor,$$

$$|y_3| \leq \left\lfloor \sqrt{\frac{\det A_3}{\det A}} \sqrt{n} \right\rfloor = \left\lfloor \sqrt{\frac{763}{30}} \sqrt{n} \right\rfloor,$$

$$|y_4| \leq \left\lfloor \sqrt{\frac{\det A_4}{\det A}} \sqrt{n} \right\rfloor = \left\lfloor \sqrt{\frac{11}{6}} \sqrt{n} \right\rfloor.$$

We remark that these are the same bounds that we obtained previously in the example. A simple computer search through these ranges for each $n \in \{1, 2, 3, 5, 6, 7, 10, 14, 15\}$ found a solution for each of the nine values of n.

n	(y_1, y_2, y_3, y_4)	n	(y_1, y_2, y_3, y_4)
1	$(-5, -2, 4, 1)$	7	$(-13, -5, 10, 3)$
2	$(-4, -2, 4, 1)$	10	$(-1, 0, 1, 0)$
3	$(-9, -4, 8, 2)$	14	$(-22, -9, 18, 5)$
5	$(-7, -3, 6, 2)$	15	$(-21, -9, 18, 5)$
6	$(-14, -6, 12, 3)$		

Thus, by the strong 15-Theorem, $G(x_1, x_2, x_3, x_4)$ is universal.

The solutions (y_1, y_2, y_3, y_4) in the table are not unique since if (y_1, y_2, y_3, y_4) is a solution so is $(-y_1, -y_2, -y_3, -y_4)$. However, if we identify the two solutions $\pm(y_1, y_2, y_3, y_4)$ the solutions given in the table for $n = 1$ and $n = 2$ are unique. For $n = 6$, with this identification, there are six solutions, namely,

$$\pm(2, 1, -2, -1), \quad \pm(4, 1, -2, -1), \quad \pm(6, 2, -4, -1),$$

$$\pm(6, 3, -6, -1), \quad \pm(12, 5, -10, -3), \quad \pm(14, 6, -12, -3).$$

The solution $\pm(14, 6, -12, -3)$ lies on the boundary of the hypercube as

$$\left\lfloor \sqrt{\frac{361}{10}}\sqrt{6} \right\rfloor = 14, \quad \left\lfloor \sqrt{\frac{19}{3}}\sqrt{6} \right\rfloor = 6, \quad \left\lfloor \sqrt{\frac{763}{30}}\sqrt{6} \right\rfloor = 12, \quad \left\lfloor \sqrt{\frac{11}{6}}\sqrt{6} \right\rfloor = 3,$$

again showing that in general the theorem is best possible.

Acknowledgment The author thanks the referees and the editor for their valuable suggestions which led to improvements in this article.

REFERENCES

1. A. A. Albert, The integers represented by sets of ternary quadratic forms, *Amer. J. Math.* **55** (1933) 274–292. [*Collected Mathematical Papers of A. Adrian Albert*. Part 2. American Mathematical Society, Providence, RI, 1993. 57–75.]
2. M. Bhargava, On the Conway–Schneeberger fifteen theorem, *Contemp. Math.* **272** (2000) 27–37.
3. M. Bhargava, J. Hanke, Universal quadratic forms and the 290-theorem, preprint.
4. J. H. Conway, *The Sensual (Quadratic) Form*. The Carus Mathematical Monographs, No. 26. Mathematical Association of America, Washington, DC, 2005.
5. J. H. Conway, Universal quadratic forms and the fifteen theorem, *Contemp. Math.* **272** (2000) 23–26.
6. L. E. Dickson, Quadratic forms which represent all integers, *Proc. Natl. Acad. Sci. USA* **12** no. 12 (1926) 756–757. [*Collected Mathematical Papers of Leonard Eugene Dickson*. Vol. III. Chelsea Publishing, New York, 1975. 443–444.]
7. L. E. Dickson, Quaternary quadratic forms representing all integers, *Amer. J. Math.* **49** no. 1 (1927) 39–56. [*Collected Mathematical Papers of Leonard Eugene Dickson*. Vol. I, Chelsea Publishing, New York, 1975. 461–478.]
8. L. E. Dickson, *Modern Elementary Theory of Numbers*. Univ. of Chicago Press, Chicago, 1939.
9. J. L. Lagrange, Démonstration d'un théorème d'arithmétique, *Nouveaux Mémoires de l'Académie royale des Sciences et Belles-Lettres de Berlin,* 1770; *Oeuvres de Lagrange, Tome Troisième, publiées pars les soins de M. J.-A. Serret, sous les auspices de son excellence le ministre de l'instruction publique, Gauthiers-Villars, Paris,* 1869, pp. 189–201.
10. J. Liouville, Sur la forme $x^2 + y^2 + 2(z^2 + t^2)$, *J. Pures Appl. Math.* **5** (1860) 269–272.
11. L. Mirsky, *An Introduction to Linear Algebra*. Oxford University Press, Oxford, 1972.
12. S. Ramanujan, On the expression of a number in the form $ax^2 + by^2 + cz^2 + du^2$, *Proc. Cambridge Philos. Soc.* **19** (1917) 11–21. [Collected Papers of Srinivasa Ramanujan, AMS Chelsea Publishing, Amer. Math. Soc., Providence, Rhode Island, 2000, pp. 169–178.]
13. W. A. Schneeberger, *Arithmetic and geometry of integral lattices*, Ph.D. dissertation, Princeton Univ., 1995.

Summary. Recent ground-breaking work of Conway, Schneeberger, Bhargava, and Hanke shows that to determine whether a given positive quadratic form F with integer coefficients represents every positive integer (and so is universal), it is only necessary to check that F represents all the integers in an explicitly given finite set S of positive integers. The set contains either nine or twenty-nine integers depending on the parity of the coefficients of the cross-product terms in F and is otherwise independent of F. In this article we show that F represents a given positive integer n if and only if $F(y_1, \ldots, y_k) = n$ for some integers $y_1, \ldots, y_k$ satisfying $|y_i| \leq \sqrt{c_i n}$, $i = 1, \ldots, k$, where the positive rational numbers c_i are explicitly given and depend only on F. Let m be the largest integer in S (in fact $m = 15$ or 290). Putting these results together we have

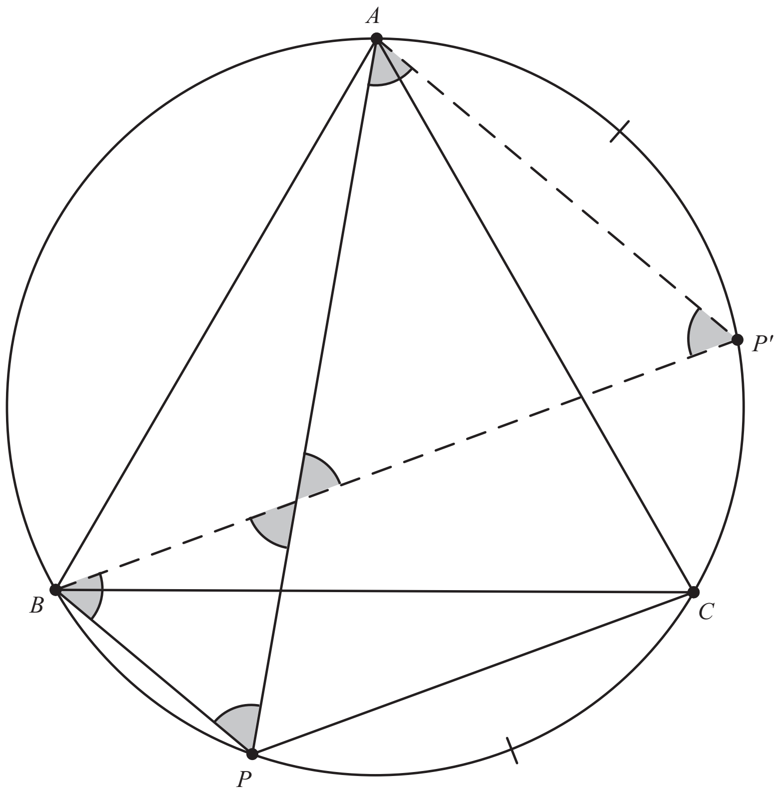
$$F \text{ is universal if and only if } S \subseteq \{F(y_1, \ldots, y_k) \mid |y_i| \leq \sqrt{c_i m}, i = 1, \ldots, k\}.$$

KENNETH S. WILLIAMS (MR Author ID: 183165) received his Ph.D. degree in mathematics from the University of Toronto in 1965. In 1979 he was awarded a D.Sc. degree from the University of Birmingham, England for his research in the theory of numbers. After 36 years as a faculty member at Carleton University, he retired in 2002 as Professor Emeritus and Distinguished Research Professor. His interests include birding with his wife, gardening, bridge, chess, and, once a week, looking after his seven year old granddaughter Isabelle Sofie Olsen (with his wife's help of course).

Proof Without Words: van Schooten's Theorem

RAYMOND VIGLIONE
Kean University
Union, NJ, 07083
rviglion@kean.edu

An equilateral triangle ABC along with its circumcircle is given, and point P is chosen on minor arc BC. Then $PA = PB + PC$.



Summary. We provide a simple visual proof of van Schooten's theorem: given an equilateral triangle ABC with circumcircle, if point P is chosen on minor arc BC, then $PA = PB + PC$.

RAYMOND VIGLIONE (MR Author ID: 704098) is a mathematics professor at Kean University who cheerfully refuses to grow up.

Math. Mag. **89** (2016) 132–132. doi:10.4169/math.mag.89.2.132.
MSC: Primary 51M04.

Intransitive Dice

BRIAN CONREY
American Institute of Mathematics
San Jose, CA
conrey@aimath.org

JAMES GABBARD
University of Southern California
jgabbard@usc.edu

KATIE GRANT
University of California, San Diego
k.grant529@gmail.com

ANDREW LIU
Live Oak High School
Morgan Hill, CA
andrewliu54@gmail.com

KENT E. MORRISON
American Institute of Mathematics
San Jose, CA
morrison@aimath.org

In 1970, Martin Gardner introduced intransitive (also called "nontransitive") dice in his *Mathematical Games* column [**3**]. The particular dice he described were invented by Bradley Efron a few years earlier. The six face values for the four dice A, B, C, and D are

$$A = (0, 0, 4, 4, 4, 4) \quad B = (3, 3, 3, 3, 3, 3)$$
$$C = (2, 2, 2, 2, 6, 6) \quad D = (1, 1, 1, 5, 5, 5)$$

The result is a paradoxical cycle of dominance in which

- A beats B with probability $2/3$
- B beats C with probability $2/3$
- C beats D with probability $2/3$
- D beats A with probability $2/3$

For example, consider rolling C and D. Of the 36 outcomes, there are 24 for which the value shown by C is greater than the value of D.

Effron's dice provide a concrete example of what was first noticed in 1959 by Steinhaus and Trybuła [**9**] in a short note (with no mention of dice) showing the existence of independent random variables X, Y, and Z such that $P(X > Y) > 1/2$, $P(Y > Z) > 1/2$, and $P(Z > X) > 1/2$. This was followed with expanded versions by Trybuła containing the details and proofs [**11, 12**]. Notably, he found equations that describe the maximal probabilities possible for an intransitive cycle of m random variables. For $m = 4$, this maximal probability is $2/3$, which means that Efron's dice are optimal in this sense.

Math. Mag. **89** (2016) 133–143. doi:10.4169/math.mag.89.2.133.
MSC: Primary 91A60, Secondary 60C05.

Starting with six-sided dice and then generalizing to n-sided dice, we focus in this article on just how prevalent intransitive dice are. Much of the work is experimental in nature, but it leads to some tantalizing conjectures about the probability that a random set of k dice, $k \geq 3$, makes an intransitive cycle as the number of sides goes to infinity. For a very restricted ensemble of n-sided dice, which we call "one-step dice," we prove the conjectures for three dice.

How rare are intransitive dice?

Both surprise and puzzlement are the universal reactions to learning about intransitive dice, and, indeed, that was the case for all of us, but once we had seen some examples, we began to wonder just how special they are. For example, suppose we pick three dice randomly and find that A beats B and B beats C. Does that make it more likely that C beats A?

Let's specify exactly what we mean by a random choice of dice. We begin with dice that are much like the standard die commonly used: the number of sides is six, the numbers on the faces come from $\{1, 2, 3, 4, 5, 6\}$, and the total is 21. We don't care how the six numbers are placed on the faces and so each die can be represented by a nondecreasing sequence $(a_1, a_2, a_3, a_4, a_5, a_6)$ of integers. Except for the standard die, there must be some repetition of the numbers on the faces. There are 32 such sequences.

$$(1,1,1,6,6,6),\ (1,1,2,5,6,6),\ (1,1,3,4,6,6),\ (1,2,2,4,6,6),$$
$$(1,2,3,3,6,6),\ (2,2,2,3,6,6),\ (1,1,3,5,5,6),\ (1,2,2,5,5,6),$$
$$(1,1,4,4,5,6),\ (1,2,3,4,5,6),\ (2,2,2,4,5,6),\ (1,3,3,3,5,6),$$
$$(2,2,3,3,5,6),\ (1,2,4,4,4,6),\ (1,3,3,4,4,6),\ (2,2,3,4,4,6),$$
$$(2,3,3,3,4,6),\ (3,3,3,3,3,6),\ (1,1,4,5,5,5),\ (1,2,3,5,5,5),$$
$$(2,2,2,5,5,5),\ (1,2,4,4,5,5),\ (1,3,3,4,5,5),\ (2,2,3,4,5,5),$$
$$(2,3,3,3,5,5),\ (1,3,4,4,4,5),\ (2,2,4,4,4,5),\ (2,3,3,4,4,5),$$
$$(3,3,3,3,4,5),\ (1,4,4,4,4,4),\ (2,3,4,4,4,4),\ (3,3,3,4,4,4)$$

We say that A **beats** B, denoted by $A \succ B$, if $P(A > B) > P(B > A)$, i.e., the probability that $A > B$ is greater than the probability that $B > A$. Here, we think of the dice as random variables with each of the components in their vector representations being equally likely. This is equivalent to saying that

$$\sum_{i,j} \operatorname{sign}(a_i - b_j) > 0.$$

We also say that A dominates B or that A is stronger than B. If it happens that $P(A > B) = P(B > A)$, we say that A and B tie or that they have equal strength.

For all choices of three dice (A, B, C), there are $32^3 = 32{,}768$ possibilities. With the aid of a computer program, we found 4417 triples such that $A \succ B$ and $B \succ C$. Then for the comparison between A and C, there were 930 ties, 1756 wins for A and 1731 wins for C. Therefore, knowing that $A \succ B$ and $B \succ C$ gives almost no information about the relative strengths of A and C. The events $A \succ C$ and $C \succ A$ are almost equally likely!

For each triple of dice, there are three pairwise comparisons to make, and for each comparison, there are three possible results: win, loss, tie. Throwing away the triples that have any ties leaves us with 13,898 triples and only eight comparison patterns. Our results show that each of the eight patterns occur with nearly the same frequency. Each of the six patterns that give a transitive triple occurs 1756 times, and each of the two patterns resulting in an intransitive triple occurs 1731 times. These total 10,386 transitive and 3512 intransitive.

Rather than look at all ordered triples (A, B, C), we get equivalent information from all subsets of three dice $\{A, B, C\}$, and this, of course, requires much less computation. Of the $\binom{32}{3} = 4960$ subsets, we find that 2627 of them contain ties. Of the remaining 2333 subsets, there are 1756 transitive sets and 577 intransitive sets. Allowing for the six permutations of each set, we get the same totals as for the ordered triples.

Proper dice

With only six sides the number of ties is significant, but what if we increase the number of sides on the dice and let that number grow? Define an n-sided die to be an n-tuple $(a_1, \ldots, a_n)$ of nondecreasing positive integers, $a_1 \leq a_2 \leq \cdots \leq a_n$. The **standard** n-sided die is $(1, 2, 3, \ldots, n)$. We define **proper** n-sided dice to be those with $1 \leq a_i \leq n$ and $\sum a_i = n(n+1)/2$. Thus, every proper die that is not the standard one has faces with repeated numbers.

Above, we listed the proper n-sided dice for $n = 6$. Here is a list for $n \leq 5$:

$n = 1$ (1)

$n = 2$ (1, 2)

$n = 3$ (1, 2, 3), (2, 2, 2)

$n = 4$ (1, 1, 4, 4), (1, 2, 3, 4), (1, 3, 3, 3), (2, 2, 2, 4), (2, 2, 3, 3)

$n = 5$ (1, 1, 3, 5, 5), (1, 1, 4, 4, 5), (1, 2, 2, 5, 5), (1, 2, 3, 4, 5),
(1, 2, 4, 4, 4), (1, 3, 3, 3, 5), (1, 3, 3, 4, 4), (2, 2, 2, 4, 5),
(2, 2, 3, 3, 5), (2, 2, 3, 4, 4), (2, 3, 3, 3, 4), (3, 3, 3, 3, 3)

The number of proper n-sided dice occurs as sequence A076822 in the Online Encyclopedia of Integer Sequences [**7**], where it is described as the number of partitions of the n-th triangular number $n(n+1)/2$ into exactly n parts, each part not exceeding n. Below are the terms of the sequence for $n \leq 27$.

1	1	15	1328980
2	1	16	4669367
3	2	17	16535154
4	5	18	58965214
5	12	19	211591218
6	32	20	763535450
7	94	21	2769176514
8	289	22	10089240974
9	910	23	36912710568
10	2934	24	135565151486
11	9686	25	499619269774
12	32540	26	1847267563742
13	110780	27	6850369296298
14	381676		

Obviously, the number of proper dice grows rapidly, and while it is not necessary to our understanding of intransitive dice, we were curious about the rate of growth. Surprisingly, the OEIS entry has nothing about the asymptotics of these partition numbers, but with some heuristics involving the central limit theorem, we were able to conjecture that the n-th term is asymptotic to

$$\frac{\sqrt{3}}{2\pi}\frac{4^n}{n^2}.$$

Eventually, we found this result proved rigorously in a 1986 paper by Takács [**10**], although it is no trivial task to make the connection. You can see the dominant power of 4 in the numbers above. A question that we have been unable to answer is whether there is some construction that gives (approximately) four proper dice with $n+1$ sides from each proper die with n sides.

Two conjectures for three dice

Arising from our computer simulations are two conjectures about random sets of three n-sided dice as $n \to \infty$.

Conjecture 1. *In the limit, the probability of any ties is* 0.

Conjecture 2. *In the limit, the probability of an intransitive set is* $1/4$.

We can state Conjecture 2 using random ordered triples rather than random sets. For three dice A, B, C, there are eight different dominance patterns when there are no ties. In the limit as $n \to \infty$, we conjecture that each of these patterns has a probability approaching $1/8$. Since two of the patterns are intransitive and six are transitive, the intransitive probability approaches $1/4$ and the transitive probability approaches $3/4$.

Although the conjecture deals with the behavior as n grows, the data for small n already show us something. For $n = 4$, among the ten sets of three distinct dice, the only intransitive set is the set $\{(1, 1, 4, 4), (1, 3, 3, 3), (2, 2, 2, 4)\}$. There is also just one transitive set, while the other eight sets have ties. Thus, the proportion of intransitive is $1/10$. For $n = 5$, there are $\binom{12}{3} = 220$ sets, and 23 of these are intransitive with a ratio of $23/220 \approx 0.105$. (There are 54 transitive sets.) For $n = 7$, the computer calculations showed the proportion of intransitive among all the sets is $19929/134044 \approx 0.149$.

A proof of Conjecture 2 appears to be difficult, and we do not know how to attack it, but in a later section, we present rigorous results about a much smaller collection of dice, the "one-step dice." Conjecture 1, on the other hand, appears to be more attainable, and here we provide a plausibility argument for it.

First, Conjecture 1, which involves three dice, holds if we can prove that the probability of a tie between two random n-sided dice goes to 0 as $n \to \infty$. That is simply because the probability of any tie is less than or equal to three times the probability of a tie between two dice. We represent a proper n-sided die by a vector $(v_1, \ldots, v_n)$ where v_i is the number of faces with the value i. There are two restrictions: $\sum v_i = n$, which means that there are n faces, and $\sum i v_i = n(n+1)/2$. Letting D_n be the set of proper n-sided dice, we see that D_n is the set of integer lattice points in the intersection of $[1, n]^n$ with the $(n-2)$-dimensional affine subspace of $\mathbf{R}^n$ defined by the two restrictions. When we roll dice v and w, there are n^2 possible outcomes. When v has the value i showing, then it is greater than $w_1 + \cdots + w_{i-1}$ of the faces of w, and it is less than $w_{i+1} + \cdots + w_n$ faces. Summing over i, we see that v and w are equally strong when they satisfy the polynomial equation

$$\sum_{i<j} v_i w_j - \sum_{i>j} v_i w_j = 0.$$

This equation defines a hypersurface in $\mathbf{R}^n \times \mathbf{R}^n$, and the set of pairs of tied dice (v, w) is the intersection of $D_n \times D_n$ with this hypersurface. Heuristically, this means that the dimension of the set of tied dice is one less than the dimension of the set of all pairs. Since the coordinates need to be integers between 0 and n, this suggests that the ratio of the number of tied pairs to the number of all pairs, which is the probability of a tie, should be approximately $1/n$. Computer simulations with 10,000 sample pairs for each size show roughly that behavior.

n	$1/n$	Ties
10	.100	.0864
20	.050	.0329
30	.033	.0190
40	.025	.0124
50	.020	.0131
60	.016	.0101
70	.014	.0078
80	.013	.0061
90	.011	.0053
100	.010	.0053

A serious difficulty in making this heuristic approach more rigorous is the fact that the both the coordinate range $[1, n]$ and the dimensions of the geometric objects (the affine subspace and the hypersurface) are growing with n. However, recent work by Cooley, Ella, Follett, Gilson, and Traldi [**1**] proves that the proportion of ties goes to 0 as $n \to \infty$ for n-sided dice with values between 1 and a fixed integer k and having a total equal to $n(k + 1)/2$.

For $n > 7$, we have estimated the probability of intransitive triples by sampling from the sets of proper dice. Our data, based on 10,000 sample triples for each of 10-sided, 20-sided, 30-sided, 40-sided, and 50-sided dice, are below. It shows the proportion with ties and the proportion that are intransitive.

n	Ties	Intransitive
10	.2219	.1933
20	.0862	.2267
30	.0557	.2357
40	.0439	.2380
50	.0306	.2448

Other ensembles

We have investigated other ensembles of dice in an effort to see whether the $1/4$ probability of being intransitive is a widespread phenomenon. Suppose we consider n-sided dice with the only restriction being that the total is $n(n + 1)/2$, thus allowing values greater than n. Let's call them "improper dice." These dice are the partitions of $n(n + 1)/2$ with exactly n parts. There are significantly more of them. For $n = 10$, there are 33,401 improper dice compared with 2934 proper dice. In a random sample of 1000 triples of these dice with 20 sides, we found 13 intransitive sets, 958 transitive sets, and 29 with one or more ties. These results are far different than for proper dice! What's the cause? You can visualize an n-sided die by looking at the plot in the plane of the point set $\{(i, a_i)\}$. The left side of Figure 1 shows the superimposed plots for ten random improper dice with 30 sides. The right side shows the same for ten random proper dice. You can see that the typical proper die is much closer to the standard die than the typical improper die.

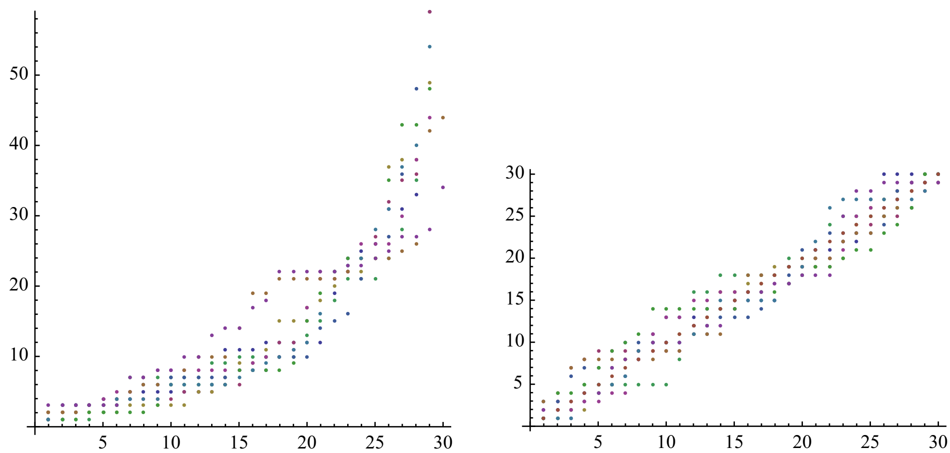


Figure 1 Left: ten random improper dice. Right: ten random proper dice.

For another model for random n-sided dice, we take n random numbers in the unit interval and sort them into increasing order. Then we rescale them, first by dividing by their total and then multiplying by $n(n+1)/2$ so that now the total is the same as for proper n-sided dice. These random dice look a lot more like the proper dice, but they still have some values greater than n. Figure 2 shows a sample of ten of them with 30 sides. We generated 1000 triples of these dice with $n = 30$ and got 130 triples with one or more ties. Of the remaining 870, there were 151 intransitive sets, giving a ratio of $151/870 \approx 0.174$. There is less intransitivity for these dice than for proper dice but still much more than for the improper dice. Random samples with more sides show the ties decreasing and the proportion of intransitive staying around 0.17 or 0.18. We do not have enough evidence to hazard a guess for the limiting value of the proportion.

Finally, if we consider n-sided dice with face numbers from 1 to n but no restriction on the total, then random samples of three such dice almost never produce ties or

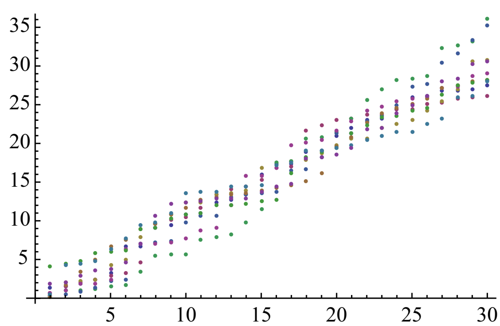


Figure 2 Ten random dice with real entries.

intransitive triples. For example, in one run of 1000 triples of 50-sided dice, there were three triples with a tie and three intransitive triples.

One-step dice

The dice that are closest to the standard die are obtained by moving a single pip from one face to another. That is, the value on one face increases by one and the value on another face decreases by one. If we define the distance between two dice A and B to be

$$\sum_i |a_i - b_i|,$$

then these dice are the minimal distance of 2 from the standard die. We call them "one-step dice," because they are one step away from the standard die in the graph whose vertices are proper dice and edges between nearest neighbors.

Let $s(a, b)$ denote the one-step die in which side a goes up by 1 and side b goes down by 1. For example, with $n = 8$,

$$s(2, 5) = (1, 3, 3, 4, 4, 6, 7, 8) \text{ and } s(5, 2) = (1, 1, 3, 4, 6, 6, 7, 8).$$

In the first, the 2 changes to 3 and the 5 changes to 4. In the second, the 5 changes to 6 and the 2 changes to 1. The die $s(a, b)$ has a repeated value of $a + 1$ and a repeated value of $b - 1$ so that $s(a, b)$ has two pairs of repeated values unless $a + 1 = b - 1$, in which case it has one value repeated three times.

Now the number of one-step dice is much smaller than the number of proper dice. We leave it to the reader to verify that the number of one-step n-sided dice is $(n - 2)^2$. With such a restricted ensemble of dice, we wondered whether we could understand the prevalence of intransitive sets more completely than for all proper dice. However, for one-step dice, ties are common. The one-step dice are not much different from the standard die, and the standard die ties all other proper dice, a fact that we'll need in the next proof. (To see that the standard die ties everyone else, use the representation of proper dice in the heuristic proof of Conjecture 1.)

Proposition 1. *As $n \to \infty$, the probability of a tie between a random pair of one-step dice goes to* 1.

Proof. We consider what happens in the comparison between two dice when we change the value on one face of one die. Suppose that with A we increase the value by one on a single face by replacing i by $i + 1$. If B has one face with the value i and one face with value $i + 1$, then in the tally of comparisons between all faces of the two dice, there is a net increase of one win for A. Similarly, if we reduce one face of A by one, say from i to $i - 1$, and if B has a one face with i and once face with $i - 1$, then A has a net increase of one loss. We first compare the standard die with $s(c, d)$, which is a tie because the standard die ties every proper die. Now we change the standard die to make it $s(a, b)$. The result is a tie as long as a, $a + 1$, b, and $b - 1$ are not repeated values for $s(c, d)$. Thus, the two dice will tie if $|a - c|, |a - d|, |b - c|, |b - d| > 2$. As n increases and the values of a, b, c, d are selected randomly, the probability that these inequalities hold approaches 1.

So, we have a major difference between one-step dice and all proper dice: As n grows, ties become more likely for one-step dice and less likely for proper dice. On the other hand, when we just look at triples of one-step dice in which there are no ties, we see the same behavior as for proper dice: Very close to one-fourth of the triples are intransitive. For $n = 10$ there are 64 one-step dice and $\binom{64}{3} = 41{,}664$ sets of three dice, of which 8086 have no ties. There are 2072 intransitive sets, a proportion of 0.256. With $n = 20$, there are 324 one-step dice, and $\binom{324}{3} = 5{,}616{,}324$ sets of size three. We randomly sampled 100,000 sets and found 3664 with no ties, 907 of them intransitive, for a proportion of 0.2475.

Next, we analyze the four scenarios in which one of the modified faces of $s(a, b)$ is close to one of the modified faces of $s(c, d)$ to find out what must hold so that $s(a, b) \succ s(c, d)$. For example, consider the possibility that a and c are close, which means $|a - c| \leq 2$. We can assume that $a \geq c$ without loss of generality. If $a = c$, there is a tie. Now consider $a = c + 1$. The first die $s(a, b) = s(c + 1, b)$ now has among its face values the sequence $c - 1, c, c + 2, c + 2$ while the second die $s(c, d)$ contains the sequence $c - 1, c + 1, c + 1, c + 2$. (The rest of the values of the two dice are not relevant.) In the 16 pairwise comparisons of these, the first die wins seven, loses six, and ties three. Therefore, $s(c + 1, b)$ beats $s(c, d)$. The other possibility is that $a = c + 2$. The die $s(a, b) = s(c + 2, b)$ has the face values $c, c + 1, c + 3, c + 3$, while $s(c, d)$ contains $c - 1, c + 1, c + 1, c + 2$. Now the 16 pairwise comparisons result in six wins for each die and four ties, and so $s(c + 2, b)$ and $s(c, d)$ are of equal strength.

By analyzing each of the other three possibilities for a or b interacting with c or d, we establish the following lemma.

Lemma 1. *In order for $s(a, b)$ to dominate $s(c, d)$, one or more of the following must hold:*

$$a = c + 1$$
$$d = a + 2$$
$$b = c$$
$$b = d + 1.$$

Proposition 2. *If A, B, C are randomly chosen one-step dice with no ties among them such that $A \succ B \succ C$, then in the limit as $n \to \infty$, the two outcomes $A \succ C$ (transitive) and $C \succ A$ (intransitive) are equally likely. Consequently, if three randomly chosen one-step dice have no ties among them, then in the limit as $n \to \infty$, the probability that they are intransitive approaches* $1/4$.

Proof. With the lemma and the help of a computer program, we can estimate the number of solutions for the two alternatives:

$$s(a, b) \succ s(c, d) \succ s(e, f) \prec s(a, b) \quad \text{(transitive)}$$
$$s(a, b) \succ s(c, d) \succ s(e, f) \succ s(a, b) \quad \text{(intransitive).}$$

From the lemma, we see that each comparison can occur in four ways. Each alternative requires three comparisons, and so there are potentially 4^3 scenarios for each. However, some of them are logically impossible; for example, in the intransitive alternative, the choices $a = c + 1$, $c = e + 1$, and $e = a + 1$ lead to the contradiction $a = a - 2$. Now each scenario is represented by a system of three linear equations in the six variables a, b, c, d, e, f. Our computer program checks each of the 64 systems to find those that have positive integer solutions corresponding to one-step dice. The result is that for each alternative 47 of the 64 are feasible.

Because of boundary effects, the scenarios do not have exactly the same number of solutions, but they each have on the order of n^3 solutions since there are three free variables. The boundary effects result in a lower order correction to the dominant n^3

term. Therefore, the number of solutions for each alternative is on the order of $47n^3$, and so in the limit, the two alternatives are equally likely.

The big conjectures

We have seen that intransitive sets of three dice are actually quite common, but what about longer cycles of intransitive dice? Do they even occur? Is there a maximal length? In 2007, Finkelstein and Thorp [**2**] gave an explicit construction of intransitive cycles of arbitrary length. For example, their construction gives an intransitive cycle of length 5

$$A_1 \succ A_2 \succ A_3 \succ A_4 \succ A_5 \succ A_1$$

with these 15-sided dice:

$$A_1 = (7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 12, 12, 12)$$
$$A_2 = (6, 6, 6, 6, 6, 6, 6, 6, 6, 11, 11, 11, 11, 11, 11)$$
$$A_3 = (5, 5, 5, 5, 5, 5, 10, 10, 10, 10, 10, 10, 10, 10, 10)$$
$$A_4 = (4, 4, 4, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9)$$
$$A_5 = (8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8).$$

For each odd integer, they exhibit an intransitive cycle of that length consisting of dice with three times as many sides. To get a cycle of even length, just construct a cycle of length one greater and delete one of the dice.

How common are intransitive cycles? With four dice, they are quite common. Here are the results from random samples of 1000 sets of four dice having 50, 100, 150, 200 sides.

n	Ties	Intransitive
50	.061	.359
100	.029	.365
150	.023	.381
200	.008	.392

It definitely looks like the probability of there being any ties goes to 0, but it's less clear what is happening to the intransitive probability. Before reading further, you might want to hazard a guess as to the limiting probability that four random dice form an intransitive cycle.

We have some far-reaching conjectures that go far beyond three or four dice. As consequences, we can conjecture the probability (in the limit) that a random set of k dice form an intransitive cycle or that they form a completely transitive set. These conjectures also imply that, for proper dice, the dominance relation exhibits no bias in favor of transitivity as the number of sides goes to infinity. We consider k random n-sided proper dice $A_1, A_2, \ldots, A_k$ for a fixed integer $k \geq 2$.

Conjecture 3. *The probability that there is a tie between any of the k dice goes to* 0 *as* $n \to \infty$.

When there are no ties between any of the dice, then there are $2^{\binom{k}{2}}$ outcomes for all the pairwise comparisons among the dice, and each of these outcomes is represented by a tournament graph on k vertices. The vertices are $A_1, A_2, \ldots, A_k$ and there is an edge from A_i to A_j if $A_i \succ A_j$. (A **tournament graph** is a complete directed graph and is so-called because it represents the results of a round robin tournament.) There are $2^{\binom{k}{2}}$ tournament graphs.

Conjecture 4. *In the limit as $n \to \infty$ all the tournament graphs with k vertices are equally probable.*

Let's apply this conjecture to the case of four dice. There are six comparisons among the pairs of dice and so there are $2^6 = 64$ different tournament graphs. How many of these graphs contain a cycle of length 4? There are six ways to cyclically arrange the four vertices. Then the remaining two edges can point in either direction. Thus, there are 24 tournament graphs that contain a 4-cycle. Therefore, the probability of an intransitive cycle should go to $24/64 = 3/8$. The experimental results are consistent with the $3/8$ conjecture.

Similar reasoning predicts that the probability of a completely transitive arrangement of four dice has a limit of $3/8$ because there are $4!$ tournament graphs that allow the vertices to be linearly ordered. There are two more symmetry classes of four vertex tournament graphs. In each, there is a 3-cycle with the fourth vertex either dominating or dominated by the vertices in the 3-cycle. There are eight tournament graphs in each of these symmetry classes. Our simulation results are consistent with the prediction that the probability of a completely transitive set is $3/8$, and for the other two classes the probabilities are each $1/8$.

Under the assumption that Conjecture 4 holds, you can predict the probability that a random set of k dice forms an intransitive cycle by finding the number of tournament graphs that contain a cycle through all the vertices, i.e., a Hamiltonian or spanning cycle. Let $C(k)$ be the number of such tournament graphs. The predicted probability is then

$$\frac{C(k)}{2^{\binom{k}{2}}}.$$

Basic information about these graphs can be found in the classic book by Moon [**5**], where it is shown that having a spanning cycle is equivalent to two other properties: **strongly connected** or **irreducible**. Let $C(k)$ be the number of tournament graphs with k vertices that have a spanning cycle. The $C(k)$ satisfy the equation

$$\sum_{j=1}^{k} C(j) 2^{\binom{k-j}{2}} = 2^{\binom{k}{2}},$$

and so they can be computed recursively. We have already seen that $C(3) = 2$ and $C(4) = 24$. Using these values and $C(1) = 1$ and $C(2) = 0$, we find that $C(5) = 544$. Thus, we expect the probability that five random dice are intransitive to approach $544/2^{10} = 17/32 \approx 0.531$ as the number of sides increases. (The sequence $C(k)$ appears in the Online Encyclopedia of Integer Sequences [**8**] as the "number of strongly connected labeled tournaments.")

Wouldn't you guess that the more dice you have the less likely it should be that they are intransitive? But what we are seeing is exactly the opposite. And, in fact, for tournament graphs Moon and Moser proved in 1962 [**6**] that as $k \to \infty$, the proportion with spanning cycles goes to 1. Already for $k = 16$, the proportion exceeds 0.999.

So we end up with our original beliefs turned on their heads. The dominance relation for proper dice not only fails to be transitive, it is almost as far from transitive as a binary relation can be. We do not know of any other natural example of a binary relation that shows this behavior. Furthermore, our intuition that intransitive dice are rare and that larger sets are even rarer is completely unfounded. They are common for three dice and almost unavoidable as the number of dice grows.

Acknowledgment We would like to express our appreciation to Byron Schmuland and Josh Zucker for their helpful discussions. Research supported in part by a grant from the National Science Foundation.

REFERENCES

1. C. Cooley, W. Ella, M. Follett, E. Gilson, L. Traldi, Tied dice. II. Some asymptotic results, *J. Combin. Math. Combin. Comput.*, **90** (2014) 241–248.
2. M. Finkelstein, E. O. Thorp, Nontransitive dice with equal means, in *Optimal Play: Mathematical Studies of Games and Gambling*. Edited by S. N. Ethier and W. R. Eadington, Institute for the Study of Gambling and Commercial Gaming, Reno, 2007. (Preprint: `www.math.uci.edu/~mfinkels/dice9.pdf`)
3. M. Gardner, The paradox of the nontransitive dice, *Scientific Amer.*, **223** (1970) 110–111.
4. M. Gardner, Nontransitive dice and other probability paradoxes, in *Wheels, Life, and Other Mathematical Amusements*. W. H. Freeman, New York, 1983.
5. J. W. Moon, *Topics on Tournaments*. Holt, Rinehart and Winston, New York, 1968, available from Project Gutenberg, `www.gutenberg.org/ebooks/42833`.
6. J. W. Moon, L. Moser, Almost all tournaments are irreducible, *Canad. Math. Bull.*, **5** (1962) 61–65.
7. *Online Encyclopedia of Integer Sequences*, A076822, `www.oeis.org`
8. *Online Encyclopedia of Integer Sequences*, A054946, `www.oeis.org`
9. H. Steinhaus, S. Trybuła, On a paradox in applied probabilities, *Bull. Acad. Polon. Sci.* **7** (1959) 67–69.
10. L. Takács, Some asymptotic formulas for lattice paths, *J. Statist. Plann. Inference*, **14** (1986) 123–142.
11. S. Trybuła, On the paradox of three random variables, *Zastos. Mat.*, **5** (1960/61) 321–332.
12. S. Trybuła, On the paradox of n random variables, *Zastos. Mat.*, **8** (1965) 143–156.

Summary. We consider n-sided dice whose face values lie between 1 and n and whose faces sum to $n(n+1)/2$. For two dice A and B, define $A \succ B$ if it is more likely for A to show a higher face than B. Suppose k such dice $A_1, \ldots, A_k$ are randomly selected. We conjecture that the probability of ties goes to 0 as n grows. We conjecture and provide some supporting evidence that—contrary to intuition—each of the $2^{\binom{k}{2}}$ assignments of $\succ$ or $\prec$ to each pair is equally likely asymptotically. For a specific example, suppose we randomly select k dice $A_1, \ldots, A_k$ and observe that $A_1 \succ A_2 \succ \cdots \succ A_k$. Then our conjecture asserts that the outcomes $A_k \succ A_1$ and $A_1 \succ A_k$ both have probability approaching $1/2$ as $n \to \infty$.

BRIAN CONREY (MR Author ID: 51070) is the Executive Director of the American Institute of Mathematics in San Jose, a position he has held since 1997 when AIM was founded. He has taught at the University of Illinois and Oklahoma State University. He received his B.S. from Santa Clara University and his Ph.D. from the University of Michigan. His research interests are in analytic number theory and random matrix theory.

JAMES GABBARD (MR Author ID: 1146507) is a student at the University of Southern California, working towards a B.S. in Mechanical Engineering.

KATIE GRANT (MR Author ID: 1146508) is in her second year at UC San Diego, working towards a degree in Mathematics and Economics. She is an intern for a private financial advisor through Edgemont Investments and is exploring different career options in finance.

SHANG-CHI ANDREW LIU (MR Author ID: 1146510) was born in Taipei, Taiwan and is a student at Live Oak High School in Morgan Hill, California. He plans to major in Political Science in college.

James, Katie, and Andrew began their research on intransitive dice while they were high school students participating in a math circle sponsored by the American Institute of Mathematics. Their work became a successful project that received an Honorable Mention for First Place in the 2013 California State Science Fair.

KENT E. MORRISON (MR Author ID: 127200) is professor emeritus at California Polytechnic State University, San Luis Obispo. Since 2009 he has been associated with the American Institute of Mathematics where he directs the AIM Open Textbook Initiative. He received his B.A. and Ph.D. degrees from the University of California, Santa Cruz. His research interests are varied, but in recent years they have primarily been in the areas of combinatorics, probability, and game theory.

ACROSS

1. John von Neumann appeared on one issued in the U.S.
6. Feline
9. Chopper blade
14. Pertaining to blood
15. The so-called "loneliest number"
16. Roots of ___
17. Borel or Picard
18. Rainier and Kilimanjaro: Abbr.
19. Attend a lecture informally
20. * ___ method, pioneered by Erdös
23. Kan. neighbor
24. Singer DeFranco
25. British buttocks
28. *Raiders of the Lost* ___
31. * Namesake of a method in linear algebra for solving diagonally dominant systems of linear equations
36. Nickname for the most recent Best Actor winner
37. Nickname for Terence Tao's current home state
39. Vertices, in a graph
40. * Method of proof that relies on the well-ordering principle of the natural numbers
44. Birds that fly in a V
45. "Top-___": high praise for a university
46. Padre's hermano
47. * Namesake of an iterative method for finding roots of a function
49. Largest employer of mathematicians in the U.S.
50. Prefix for -tonic or -nomial
51. ___ Zeppelin
53. Tofu base
55. * ___ method, for approximating solutions to boundary values problems for partial differential equations
62. Retch
63. College, to a Briton
64. Poet who won the 1948 Nobel Prize in Literature
66. Servant to the Addams Family
67. Pro's opposite
68. Actresses Gershon and Rodriguez
69. Acyclic connected graphs
70. Ending with journal or legal
71. + end on a battery

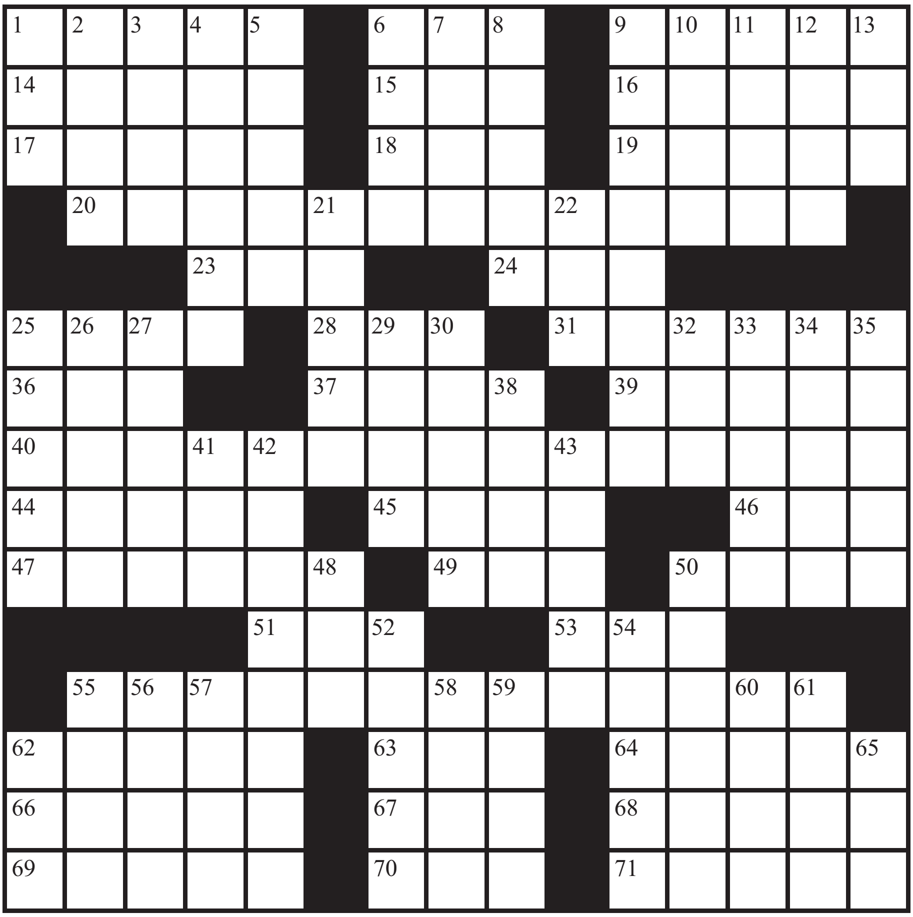
DOWN

1. "That's all ___ wrote."
2. Office fill-in
3. Mathematical historian Aczel who passed away in November 2015
4. Vivienne ___-Mayes, professor at Baylor who was the fifth African-American woman to earn a Ph.D. in mathematics in the U.S.
5. Military academy freshman
6. ___ space: topological space that's path connected but not locally path connected
7. Pro's opposite
8. Auto brand named for a scientist
9. Chebyshev, Lobachevsky, and Perelman
10. "I'm working ___!"
11. South American monkey
12. Pertaining to the ear
13. Rembrandt van ___
21. Calculators with beads
22. Another term for 1-to-1: Abbr.
25. Set straight
26. Actress Zellweger
27. *Never* ___: 1959 film starring Steve McQueen and Frank Sinatra
29. 80s hair metal band with the single "Round and Round"
30. Mathematician with a bottle?
32. Type theory created by Thierry Coquand that serves as the basis for Coq
33. Beethoven's "___ Joy"
34. West African republic
35. "How much ___ much?"
38. Bad day for Caesar
41. Suffix with algebra
42. Ancient stone implements
43. Clear the chalkboard
48. Topological object that generalizes the notion of a sequence
50. Fatty white substance that surrounds nerve axons
52. 40-40, in tennis
54. Letter that denotes the set of all finite ordinals
55. Smallest composite number
56. Philosopher of science Lakatos, who wrote *Proofs and Refutations*
57. Informal term for "well-behaved" used to describe a function, perhaps
58. Baseball Hall-of-Famer Slaughter
59. Graph of $y = ax + b$
60. El ___ (seasonal phenomenon)
61. Relative of a frog
62. Acronym for an apparatus at a Chilean observatory, the most productive astronomic facility on Earth
65. Inits. of 64-across

Math. Mag. **89** (2016) 144–146. doi:10.4169/math.mag.89.2.144.

Mathematical Methods

BRENDAN SULLIVAN
Emmanuel College
Boston, MA
sullivanb@emmanuel.edu



Clues start at left, on page 144. The Solution is on page 102.

Extra copies of the puzzle can be found at the MAGAZINE's website, `www.maa.org/mathmag/supplements`.

Crossword Puzzle Creators

If you are interested in submitting a mathematically themed crossword puzzle for possible inclusion in MATHEMATICS MAGAZINE, please contact the editor at mathmag@maa.org.

Infinity

He lectured on the mysteries of space,
the mathematical wonders of dark holes,
the numerical paths to truth and beauty.

Defined infinity as a number greater
than any quantity, mysteriously uncountable
such mythology teased me for years.

While I pondered that mystery for decades,
I pored through tomes in the libraries
of the ancients and moderns,
Egyptians, great Greeks and Romans

until I found the truth in the words
of one William Shakespeare, who said,
“Love’s not Time’s fool” Oh no, Love
will last “even to the edge of doom!”

Therefore, Professor, I have discovered
what is hiding behind the magical word,
“Infinity!” It is pronounced by an **infinite**
number of poets, defined in one word - LOVE.

— Del Corey, Emeritus
Macomb Community College
delmcorey@yahoo.com

Divergence of the Harmonic Series

The summation as k goes to infinity,
For 1 over k is solved brilliantly,
It keeps adding on,
Diverges and gone,
Its existence is proof of divinity.

— Clayton Arundel, undergraduate student, Gordon College
Communicated by Karl-Dieter Crisman, Gordon College

PROBLEMS

Correction

There was an error in the statement of Problem 1989 from the February 2016 issue. The corrected problem appears at the end of the Problems section of this issue on page 154.

Proposals

To be considered for publication, solutions should be received by September 1, 2016.

1991. *Proposed by George Stoica, Saint John, New Brunswick, Canada.*

Evaluate

$$\lim_{n\to\infty}\sum_{k=1}^{n-1}\frac{n^2}{k^2(n-k)^2}.$$

1992. *Proposed by Oniciuc Gheorghe, Botosani, Romania.*

(a) Show that there is no function $f : \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to a} |f(x)| = \infty$ for every $a \in \mathbb{Q}$.

(b) Show that there exists a function $f : \mathbb{R} \to [-\infty, +\infty]$ such that f is finite almost everywhere and $\lim_{x\to a} f(x) = +\infty$ for every $a \in \mathbb{Q}$.

Math. Mag. **89** (2016) 147–154. doi:10.4169/math.mag.89.2.147.

We invite readers to submit problems believed to be new and 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.

Effective immediately, authors of proposals and solutions should send their contributions using the Magazine's web submissions system at http://mathematicsmagazine.submittable.com. *More detailed instructions are available there. We hope that this online system will help streamline our editorial team's workflow while still proving accessible and convenient to longtime readers and contributors. We encourage submissions in PDF format, ideally accompanied by LaTeX source. General inquiries to the editors should be sent to* mathmagproblems@maa.org.

1993. *Proposed by Kimberly D. Apple, Columbus State University, GA.*

Each face of an icosahedron is colored blue or white in such a way that any blue face is adjacent to no more than two other blue faces. What is the maximum number of blue faces? (Two faces are considered adjacent if they share an edge.)

1994. *Proposed by Donald E. Knuth, Computer Science Department, Stanford University, CA.*

Let $X_0 = 0$, and suppose X_{n+1} is equally likely to be either $X_n + 1$ or $X_n - 2$. What is the probability, p_m, that $X_n \leq m$ for all $n \geq 0$?

1995. *Proposed by Michel Bataille, Rouen, France.*

Let $a_1, a_2, \ldots, a_n$ be distinct positive real numbers ($n > 2$). For $j = 1, 2, \ldots, n$, let $s_j = \sum_{i \neq j} a_i$. Consider the $n \times n$ matrix $M = (m_{i,j})$ defined by $m_{i,i} = 0$ and $m_{i,j} = a_i / s_j$ for $1 \leq i < j \leq n$. Show that there exist column vectors X and Y such that

$$X^T Y = 1 \text{ and } Y X^T = \lim_{k \to \infty} M^k.$$

Quickies

1059. *Proposed by Wong Fook Sung, Temasek Polytechnic, Singapore.*

Find a real number $\alpha > 1$ satisfying

$$\int_0^\infty \frac{1}{(1 + x^\alpha)^\alpha} \, dx = 1.$$

1060. *Proposed by Cezar Lupu, University of Pittsburgh, Pittsburgh, PA.*

Show that there exist no complex $n \times n$ matrices A, B such that A is nonzero and

$$A^* = AB - BA,$$

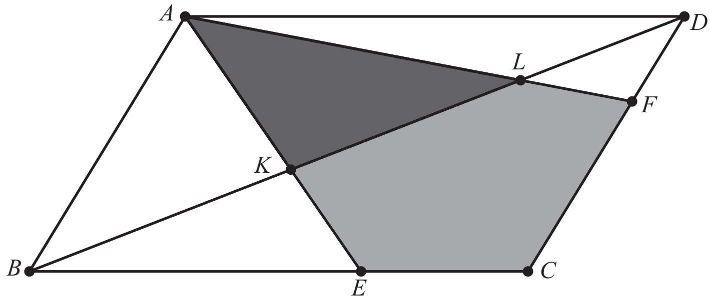
where A^* is the transpose conjugate of A.

Solutions

An area inequality in parallelograms **February 2015**

1961. *Proposed by George Apostolopoulos, Messolonghi, Greece.*

Let $ABCD$ be a parallelogram and E and F points on the sides $\overline{BC}$ and $\overline{CD}$, respectively, such that $BE \cdot FD = EC \cdot CF$. The segments $\overline{AE}$ and $\overline{AF}$ meet the diagonal $\overline{BD}$ at the points K and L, respectively. Prove that Area$(KECFL) \leq 2 \cdot$ Area(AKL).



Solution by Adnan Ali (student), A.E.C.S-4, Mumbai, India.
We use the notation $[PQR \ldots X]$ to denote the area of a convex polygon with vertices $P, Q, R, \ldots, X$. When $EC = FD = 0$, clearly $[KECFL] = [AKL] < 2[AKL]$; otherwise, let $k = BE/EC = CF/FD \geq 0$. Then

$$\frac{[ABE]}{[ABC]} = \frac{k}{k+1} \quad \text{and} \quad \frac{[AFD]}{[ACD]} = \frac{1}{k+1}.$$

On the other hand, we have $[ABC] = [ACD] = [ABCD]/2$, so $[ABE] + [AFD] = [ABC] = [ABCD]/2 = [ABD]$. Clearly,

$$[AKL] = [BEK] + [FDL].$$

Since $[ABD] = [BCD]$, we also have

$$[KECFL] = [ABK] + [ALD].$$

Thus,

$$\frac{[KECFL]}{[AKL]} = \frac{[ABK] + [ALD]}{[AKL]} = \frac{BK + LD}{KL} = \frac{BD - KL}{KL} = \left(\frac{KL}{BD}\right)^{-1} - 1. \qquad (1)$$

Next, from the similarities $\triangle ABL \sim \triangle FDL$ and $\triangle AKD \sim \triangle EKB$, we get

$$\frac{BL}{BD} = \frac{BL}{BL + LD} = \frac{AB}{AB + FD} = \frac{CF + FD}{CF + FD + FD} = \frac{k+1}{k+2},$$
$$\frac{BK}{BD} = \frac{BK}{BK + KD} = \frac{BE}{BE + AD} = \frac{BE}{BE + BE + EC} = \frac{k}{2k+1}.$$

Subtracting the second ratio above from the first:

$$\frac{KL}{BD} = \frac{BL}{BD} - \frac{BK}{BD} = \frac{k+1}{k+2} - \frac{k}{2k+1} = \frac{k^2 + k + 1}{(k+2)(2k+1)}.$$

For all $t \in \mathbb{R}$, we have $3(t^2 + t + 1) - (t + 2)(2t + 1) = (t - 1)^2 \geq 0$ with equality only if $t = 1$. Since $k \geq 0$, it follows that $KL/BD \geq 1/3$ and, from (1):

$$\frac{[KECFL]}{[AKL]} \leq \left(\frac{1}{3}\right)^{-1} - 1 = 2,$$

with equality only if $BE/EC = CF/FD = 1$, namely when E is the midpoint of $\overline{BC}$ and F the midpoint of $\overline{CD}$.

Also solved by Herb Bailey, Michel Bataille (France), Robert Calcaterra, Sayok Chakravarty, Bill Cowieson, Tim Cross, Robert L. Doucette, Dmitry Fleischman, Michael Goldenberg & Mark Kaplan, GWstat Problem Solving Group, Wei Kailai, Elias Lampakis (Greece), Kee-Wai-Lau (Hong Kong, China), Graham Lord, Peter McPolin, Titu Zvonaru & Neculai Stanciu (Romania), Dimitrios Pispinis (Saudi Arabia), Francisco Perdomo & Ángel Plaza (Spain), Joel Schlosberg, Edward Schmeichel, Randy K. Schwartz, Michael Vowe (Switzerland), and the proposer.

A limit of a series of integrals of $e^{-x^2/2}$ **February 2015**

1962. *Proposed by Timothy Hall, PQI Consulting, Cambridge, MA.*

Evaluate

$$\lim_{r\to 0^+} r^2 \sum_{k=1}^{\infty} k^2 \int_{(k-\frac{1}{2})r}^{(k+\frac{1}{2})r} e^{-x^2/2}dx.$$

Solution by Robert L. Doucette, McNeese State University, Lake Charles, LA.
For $x \in [(k-1/2)r, (k+1/2)r]$ we have $0 < x - r/2 \le kr \le x + r/2$, so

$$\int_{(k-\frac{1}{2})r}^{(k+\frac{1}{2})r} \left(x - \frac{r}{2}\right)^2 e^{-x^2/2}dx \le \int_{(k-\frac{1}{2})r}^{(k+\frac{1}{2})r} (kr)^2 e^{-x^2/2}dx \le \int_{(k-\frac{1}{2})r}^{(k+\frac{1}{2})r} \left(x + \frac{r}{2}\right)^2 e^{-x^2/2}dx.$$

Summing over $k \ge 1$, we get

$$\int_{\frac{1}{2}r}^{\infty} \left(x - \frac{r}{2}\right)^2 e^{-x^2/2}dx \le \sum_{k=1}^{\infty} \int_{(k-\frac{1}{2})r}^{(k+\frac{1}{2})r} (kr)^2 e^{-x^2/2}dx \le \int_{\frac{1}{2}r}^{\infty} \left(x + \frac{r}{2}\right)^2 e^{-x^2/2}dx. \quad (1)$$

Since $\int_0^\infty x^a e^{-x^2/2}dx$ converges for all $a > -1$, we have

$$\begin{aligned}\lim_{r\to 0^+} \int_{\frac{1}{2}r}^{\infty} \left(x - \frac{r}{2}\right)^2 e^{-x^2/2}dx &= \lim_{r\to 0^+} \int_{\frac{1}{2}r}^{\infty} x^2 e^{-x^2/2}dx \\ &\quad - \lim_{r\to 0^+} r \int_{\frac{1}{2}r}^{\infty} x e^{-x^2/2}dx + \lim_{r\to 0^+} \frac{r^2}{4} \int_{\frac{1}{2}r}^{\infty} e^{-x^2/2}dx \\ &= \int_0^\infty x^2 e^{-x^2/2}dx, \qquad (2)\end{aligned}$$

and similarly,

$$\lim_{r\to 0^+} \int_{\frac{1}{2}r}^{\infty} \left(x + \frac{r}{2}\right)^2 e^{-x^2/2}dx = \int_0^\infty x^2 e^{-x^2/2}dx. \quad (3)$$

Using the standard result

$$\int_0^\infty e^{-x^2/2}dx = \sqrt{2\pi}/2,$$

along with integration by parts, we have

$$\lim_{r\to 0^+} r^2 \sum_{k=1}^{\infty} k^2 \int_{(k-\frac{1}{2})r}^{(k+\frac{1}{2})r} e^{-x^2/2}dx = \int_0^\infty x^2 e^{-x^2/2}dx = \sqrt{2\pi}/2,$$

from (1), (2), (3) and the squeeze theorem.

Also solved by Robert Agnew, Robert Calcaterra, Hongwei Chen, Bill Cowieson, Eugene A. Herman, Michael Nathanson, Edward Schmeichel, and the proposer. There were five incomplete or incorrect solutions.

A geometric analogue of the AM–GM inequality **February 2015**

1963. *Proposed by D. M. Bătineţu-Giurgiu, Matei Basarab National College, Bucharest, Romania and Neculai Stanciu, George Emil Palade Secondary School, Buzău, Romania.*

Consider an arbitrary triangle. Let R, r, and s denote the circumradius, the inradius, and the semiperimeter of the triangle, respectively. Prove that if x and y are positive real numbers, then

$$2s(s^2 - 6Rr - 3r^2)x + (s^2 + 4Rr + r^2)y \geq 8\sqrt[4]{12x(Rrsy)^3}.$$

Solution by Habib Y. Far, Lone Star College–Montgomery, Conroe, TX.
Let a, b, c be the side lengths of the triangle, so the semiperimeter of the triangle is $s = (a + b + c)/2$, its area $rs = \sqrt{s(s-a)(s-b)(s-c)}$, and its circumradius $R = abc/(4rs)$. Thus,

$$\begin{aligned} 2s(s^2 - 6Rr - 3r^2) = 2s^3 - 12Rrs - 6r^2s &= 2s^3 - 3abc - 6(s-a)(s-b)(s-c) \\ &= a^3 + b^3 + c^3 \\ &\geq 3abc, \end{aligned}$$

by the AM–GM inequality, and

$$\begin{aligned} s^2 + 4Rr + r^2 &= \frac{s^3 + 4Rrs + r^2s}{s} = \frac{s^3 + abc + (s-a)(s-b)(s-c)}{s} \\ &= ab + bc + ca. \end{aligned}$$

By another application of the AM–GM inequality, we obtain

$$\begin{aligned} 2s(s^2 - 6Rr - 3r^2)x + (s^2 + 4Rr + r^2)y &\geq 3abcx + aby + bcy + cay \\ &\geq 4\sqrt[4]{3xa^3b^3c^3y^3} = 4\sqrt[4]{3x(4Rrsy)^3} \\ &= 8\sqrt[4]{12x(Rrsy)^3}. \end{aligned}$$

Also solved by Adnan Ali (India), Michel Bataille (France), Robert Calcaterra, Bill Cowieson, Elias Lampakis (Greece), Kee-Wai- Lau (Hong Kong, China), Rituraj Nandan, Michael Vowe (Switzerland), and the proposer.

A matrix with a geometric sequence of eigenvalues **February 2015**

1964. *Proposed by Branko Ćurgus, Western Washington University, Bellingham, WA.*

Let n be a positive integer and $a_1, a_2, \ldots, a_n$ be distinct real numbers. Consider the $n \times n$ Vandermonde matrix defined as

$$V(a_1, a_2, \ldots, a_n) = \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{pmatrix}.$$

Prove that, for an arbitrary nonzero real number x, the eigenvalues of the matrix

$$V(xa_1, xa_2, \ldots, xa_n)(V(a_1, a_2, \ldots, a_n))^{-1}$$

are $1, x, x^2, \ldots, x^{n-1}$.

Solution by Bill Cowieson, Fullerton College, CA.
We will use the notations $\mathbf{a}$ for $(a_1, a_2, \ldots, a_n)$ and $x\mathbf{a}$ for $(xa_1, xa_2, \ldots, xa_n)$. Let X be the $n \times n$ diagonal matrix with diagonal entries $1, x, x^2, \ldots, x^{n-1}$. Clearly,

$$V(x\mathbf{a}) = V(\mathbf{a})X.$$

By the classical evaluation of the Vandermonde determinant, we have

$$\det(V(\mathbf{a})) = \prod_{1 \le i < j \le n} (a_j - a_i) \neq 0,$$

since $a_1, a_2, \ldots, a_n$ are distinct. It follows that $V(\mathbf{a})$ is invertible, and

$$V(x\mathbf{a})V(\mathbf{a})^{-1} = V(\mathbf{a})XV(\mathbf{a})^{-1}.$$

Hence, the matrix $V(x\mathbf{a})V(\mathbf{a})^{-1}$ is similar to X, and it has the same eigenvalues $1, x, x^2, \ldots, x^{n-1}$ as X.

Also solved by Robert A. Agnew, Michel Bataille (France), Brian Bradie, Paul Budney, Bruce Burdick, Robert Calcaterra, Hongwei Chen, Robert L. Doucette, Dmitry Fleischman, Michael Goldenberg & Mark Kaplan, Eugene Herman, Elias Lampakis (Greece), Michael Nathanson, Northwestern University Math Problem Solving Group, Nora Thornber, Moubinool Omarjee (France), Dimitrios Pispinis (Saudi Arabia), Joel Schlosberg, Randy K. Schwartz, Haohao Wang & Jerzy Wojdylo, and the proposer. There was one incomplete or incorrect solution.

A degree-9 inhomogenous inequality in three variables **February 2014**

1965. *Proposed by Elias Lampakis, Kiparissia, Greece.*

Let x, y, and z be real numbers such that $x > 1$, $y > 1$, $z > 1$, and $x + y + z = xyz$. Prove that,

$$\sum_{\text{cyc}} (3x^5 - 10x^3 + 3x)(1 - 3y^2) \le xyz(3 - x^2)(3 - y^2)(3 - z^2),$$

where the sum runs cyclically over the variables x, y, and z.

Solution by the proposer.

For any real t with $t \neq \pm 1/\sqrt{3}$, let $R(t) = t(3-t^2)/(1-3t^2)$. Note that $R(\tan\theta) = \tan(3\theta)$ whenever $\tan\theta$ and $\tan(3\theta)$ are both defined, i. e., whenever $\tan\theta \neq \pm 1/\sqrt{3}$. For $x, y, z > 1$, let $a = \tan^{-1} x$, $b = \tan^{-1} y$, and $c = \tan^{-1} z$. We have,

$$
\begin{aligned}
x + y + z = xyz \Rightarrow & \frac{x+y}{1-xy} = -z \\
& \Rightarrow \tan(a+b) = -\tan c \\
& \Rightarrow \tan(a+b+c) = 0 \\
& \Rightarrow a+b+c = \pi, \qquad \text{since } a, b, c \in (\pi/4, \pi/2), \\
& \Rightarrow 3a + 3b + 3c = 3\pi \\
& \Rightarrow \tan(3a+3b) = \tan(3\pi - 3c) = -\tan(3c) \in (-\infty, 1) \\
& \qquad \text{since } 3\pi/2 < 3a + 3b = 3\pi - 3c < 9\pi/4 \\
& \Rightarrow \frac{R(x) + R(y)}{1 - R(x)R(y)} = -R(z) \\
& \Rightarrow R(x) + R(y) + R(z) = R(x)R(y)R(z).
\end{aligned}
$$

From the last equation above, we get

$$\sum_{\text{cyc}} \frac{x(3-x^2)}{1-3x^2} = \prod_{\text{cyc}} \frac{x(3-x^2)}{1-3x^2} = \prod_{\text{cyc}} \frac{x(3-x^2)}{1-3y^2}. \tag{1}$$

The functions $f(t) = t(3-t^2)$ and $g(t) = 1/(1-3t^2)$ are, respectively, strictly decreasing and strictly increasing in $(1, +\infty)$. Thus, the relative ordering of the quantities $f(x)$, $f(y)$, $f(z)$ is the reverse of that of the quantities $g(x)$, $g(y)$, $g(z)$, so the rearrangement inequality gives

$$\sum_{\text{cyc}} \frac{x(3-x^2)}{1-3z^2} = \sum_{\text{cyc}} f(x)g(z) \geq \sum_{\text{cyc}} f(x)g(x) = \sum_{\text{cyc}} \frac{x(3-x^2)}{1-3x^2}. \tag{2}$$

Combining (1), (2), and multiplying by the negative quantity $(1-3x^2)(1-3y^2)(1-3z^2)$, we obtain:

$$
\begin{aligned}
\sum_{\text{cyc}} (3x^5 - 10x^3 + 3x)(1-3y^2) &= \sum_{\text{cyc}} x(3-x^2)(1-3x^2)(1-3y^2) \\
&\leq \prod_{\text{cyc}} x(3-x^2) = xyz(3-x^2)(3-y^2)(3-z^2).
\end{aligned}
$$

There was one incomplete or incorrect solution.

Answers

Solutions to the Quickies from page 148.

A1059. We change the variable to $u = 1/(1 + x^\alpha)$, so the integral becomes

$$I_\alpha := \int_0^\infty \frac{1}{(1+x^\alpha)^\alpha}\,dx = \frac{1}{\alpha}\int_0^1 u^{\alpha-\frac{1}{\alpha}-1}(1-u)^{\frac{1}{\alpha}-1}du = \frac{1}{\alpha}B\left(\alpha-\frac{1}{\alpha},\frac{1}{\alpha}\right),$$

where, for $x, y > 0$, $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ is the Beta function (Euler's integral of the first kind). Using the well-known evaluation $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ of the Beta function in terms of the Gamma function $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt$, plus the functional equation $\Gamma(x + 1) = x\Gamma(x)$ (valid for $x > 0$), we have

$$I_\alpha = \frac{1}{\alpha}\cdot\frac{\Gamma(\frac{1}{\alpha})\Gamma(\alpha-\frac{1}{\alpha})}{\Gamma(\alpha)} = \frac{\Gamma(\frac{1}{\alpha}+1)\Gamma(\alpha-\frac{1}{\alpha})}{\Gamma(\alpha)}.$$

It is clear that $\alpha = \phi = (1+\sqrt{5})/2$ (the golden ratio) is a solution to the equation $I_\alpha = 1$ since we have $\alpha = \alpha^{-1} + 1$, hence $I_\alpha = \Gamma(\alpha - \alpha^{-1}) = \Gamma(1) = 1$.

Editor's Note. One can justify differentiation under the integral sign to show that $dI_\alpha/d\alpha < 0$ for $\alpha > 1$, so I_α is strictly decreasing as a function of α, and $\alpha = \phi$ is the unique solution to $I_\alpha = 1$ in $(1, +\infty)$.

A1060. Assume that matrices A, B satisfying the equation existed, with A nonzero. Multiplying by A, we obtain $A^2B - ABA = AA^*$. This implies that $\mathrm{tr}(AA^*) = \mathrm{tr}(A^2B) - \mathrm{tr}(ABA) = 0$ by linearity of the trace and the identity $\mathrm{tr}(XY) = \mathrm{tr}(YX)$. On the other hand, $\mathrm{tr}(AA^*) = \sum_{i,j}|A_{ij}|^2$ can only be zero if A is the zero matrix: a contradiction.

Correction to Problem 1989

1989. *Proposed by Michel Bataille, Rouen, France.*

Let n be a positive integer and let a a positive real number. Define

$$H_n = \sum_{k=1}^{n}\frac{1}{k} \qquad \text{and} \qquad S_n(a) = \sum_{k=1}^{n}\binom{n}{k}a^k H_k.$$

Evaluate

$$\lim_{n\to\infty}\left(\frac{S_n(a)}{(a+1)^n} - \ln n\right).$$

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.

van der Veen, Roland, and Jan van de Craats, *The Riemann Hypothesis: A Million Dollar Problem*, MAA, 2016; xi + 144 pp, $45 (member: $33.75) (P). ISBN 978-0-88385-650-5.

Mazur, Barry,and William Stein, *Prime Numbers and the Riemann Hypothesis*, Cambridge University Press, 2016; 154 pp, $59.99, $24.99(P), $20(Kindle). http://wstein.org/rh/rh.pdf. ISNB 978-1107-10192-0, 978-1107-49943-0.

If you are fortunate, the Riemann hypothesis will be resolved during your lifetime; if you are really fortunate, you will be the one to settle it. The book by van der Veen and van de Craats grew out of a four-week Web course for high school students, with each chapter corresponding to a week. Much of the content is in the form of exercises (solutions take up one-third of the book). Computer support is provided by code given in Wolfram Alpha and in Sage. The book presumes that the reader is comfortable with calculus. The second book, by Mazur and Stein, is much less demanding in its Part I, but after that, it goes deeper than the first book, as it introduces distributions, Fourier transforms, and the Riemann spectrum. There are no exercises. The two books complement each other nicely. They represent an important part of a level of literacy in contemporary mathematics that every mathematics major should attain.

Moskovitz, Clara, Elegant equations, *Scientific American* 314 (1) (January 2016) 70–73.

The Concinnitas Project. http://www.concinnitasproject.org.

Mathematician Dan Rockmore (Dartmouth) asked 10 famous mathematicians and physicists to write out what they felt was the "most beautiful mathematical expression"; Rockmore had the results turned into aquatints, which will now tour various galleries. Your acquaintances who subscribe to *Scientific American* may ask your opinion about these—or about your own favorite expression. Plan on expounding to them on Newton's Method (chosen by Stephen Smale) or perhaps on P versus NP (Richard M. Karp) or maybe Conservation Laws (Peter Lax) (the latter two works are only at the website); the other expressions are deep and hard to explain. (The word *concinnitas* was used by Renaissance scholar Leon Battista Alberti for "the balance of elements necessary for beautiful art"; from the Latin we get the little-used English *concinnity*, used largely to refer to harmony of parts in a literary work.)

Mazur, Joseph, *Fluke: The Math and Myth of Coincidence*, Basic Books, 2016; viii + 259 pp, $26.99(P). ISBN 978-0-465-06095-5.

Numerous popular books give the frequencies of various rare and not-so-rare events (example of the latter: a person winning a multimillion dollar lottery jackpot more than once). This book distinguishes itself by being more philosophical. It starts with 10 stories of coincidences; continues on to consider relevant probability (with scarcely any equations or notation); then revisits the stories with potential explanations; and finally includes essays about DNA evidence, the chance discovery of X-rays, "flukes" in finance, extrasensory perception (ESP), and coincidences in literature.

Math. Mag. **89** (2016) 155–156. doi:10.4169/math.mag.89.2.155.

Ash, Avner, and Robert Gross, *Summing It Up: From One Plus One to Modern Number Theory*, Princeton University Press, 2016; xv + 229 pp, $27.95. ISBN 978-0-691-17019-0.

This is a remarkable book, its depth belied by the title and subtitle (which perhaps were forced on the authors). It is the third book in a trilogy, following the authors' *Fearless Symmetry* (2006) (about diophantine equations) and *Elliptic Tales* (2012) (about elliptic curves); the three books can be read independently. This book is devoted to explaining modular forms. The authors inaccurately claim that the reader does not require mathematical knowledge beyond calculus (plus complex numbers) but go on to say "but it gets rather intricate." *Yes, it does, but it is done well.* The first part deals with sums of powers of integers (theorem of Lagrange for four squares, Waring's problem on higher powers). The second part treats infinite series, analytic continuation, infinite products, and generating functions (particularly for partitions). The third part involves the group of motions of the hyperbolic plane, modular forms, more on partitions and sums of squares, Hecke operations, L-functions, and Galois representations. This is a book that explains the big picture about how modular forms can be used to solve problems—it offers the kind of balance of perspective and detail that is so rare in mathematical exposition. It gets "intricate" fast, but it also includes some concrete numerical calculations.

Alsina, Claudi and Roger B. Nelsen, *A Mathematical Space Odyssey: Solid Geometry in the 21st Century*, MAA, 2015; xiv + 272 pp, $55 ($44 for members). ISBN 978-0-88385-358-0.

Solid geometry has not survived the waves of curricular change in high school mathematics (and analytic geometry has disappeared from calculus books). This book, livelier than any of the old solid geometry material, begins with 10 examples of ideas to follow, including counting blocks, dissections, finding the radius of a ball, coloring countries in space, and imagining a solid from two-dimensional projections. There are "Challenges" (exercises), with solutions, and figures and photos abound in this absorbing book.

Caudle, Kyle, and Erica Daniels, Did the Gamemakers fix the lottery in the Hunger Games?, *Teaching Statistics* 37 (2) (2014) 37–40.

The last of the *Hunger Games* films (adapted from books by Suzanne Collins) was released in December; as an element of popular culture, the Hunger Games will be current for the next few entering classes of college students. This article compares data from the first book in the series with numbers expected under the rules for entries in the Hunger Games lottery. Authors Caudel and Daniels suggest classroom activities for comparing data with theoretical distributions, use the standard chi-square goodness-of-fit test, and compare the results of computer simulations.

Nahin, Paul J., *In Praise of Simple Physics: The Science and Mathematics behind Everyday Questions*, Princeton University Press, 2016; xxiii + 238 pp, $29.95. ISBN 978-0-691-16693-3.

Despite its title and subtitle, this is largely a book of "back-of-the-envelope problems"—sometimes requiring a large envelope and the aid of a calculator but mainly requiring analytical thinking about simple equations in mechanics. Topics include the driver's dilemma at a yellow traffic light, the burden of walking a ladder upright, shooting bullets uphill, and measuring gravity in your garage. A postscript addresses a criticism by the writer of the foreword, who lamented the lack of dimensional analysis.

Shell-Gellasch, Amy, and J.B. Thoo, *Algebra in Context: Introductory Algebra from Origins to Applications*, Johns Hopkins University Press, 2015; xi + 536 pp, $99.50. ISBN 978-1-4214-1728-8.

This book is a text/resource book that uses history to teach algebra plus other mathematics. It treats numeration systems, algorithms for arithmetic (including finding square roots), cardinality and number theory, and solving equations (including logarithms)—all in the context of their historical and cultural settings. The book is not for an Algebra I course (it assumes familiarity with elementary algebra) nor is it developmental toward calculus. It is a "gateway to appreciation" of mathematics; students will already be comfortably familiar with some but not all of the mathematics. The book is particularly suitable for mathematics education students but also for a "general education" course in mathematics. [Disclosure: Part of the book was written and class-tested while one of the authors was a colleague of mine at Beloit College.]

MAA100
MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth St., NW · Washington, DC 20036

CONTENTS

ARTICLES

PROBLEMS AND SOLUTIONS

REVIEWS