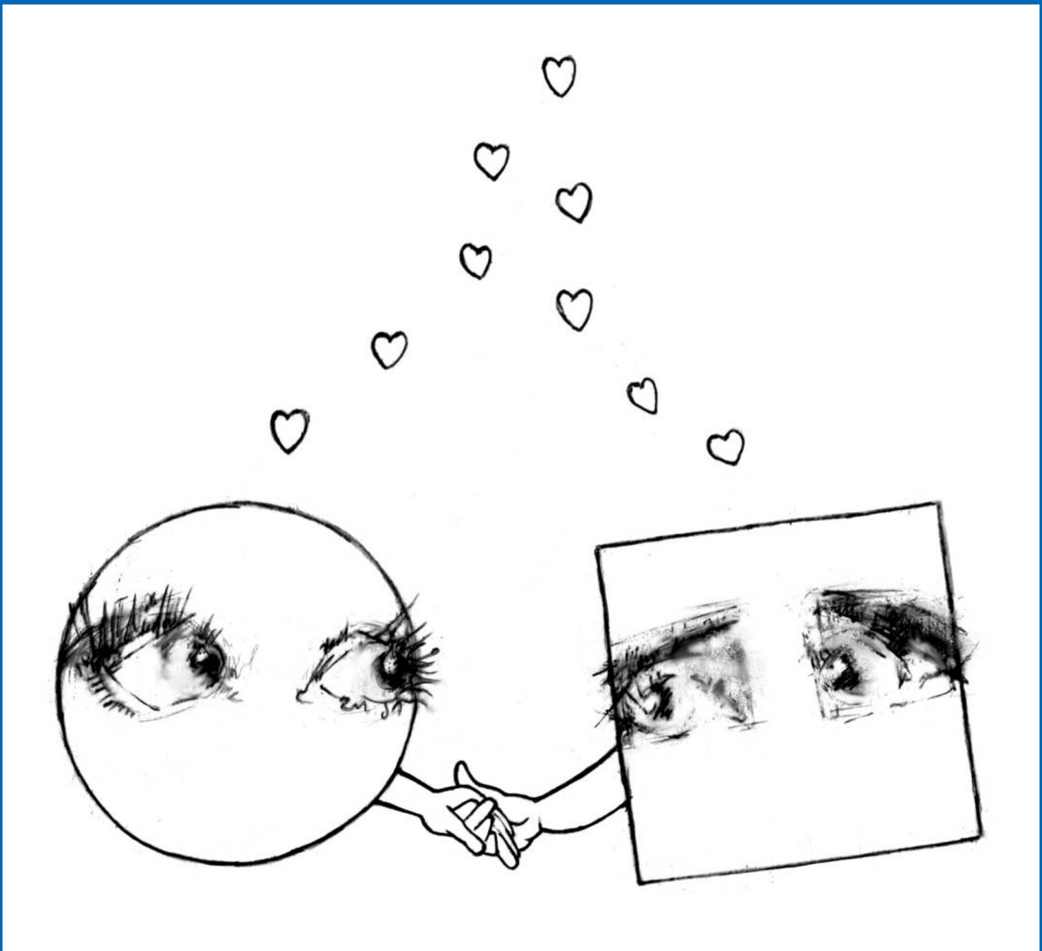
Vol. 84, No. 4, October 2011

MAA

MATHEMATICS MAGAZINE



Circle and Square

- Conway's M13 Puzzle
- Squigonometry
- Referendum Elections and Separable Preferences
- The Olympiads: USAMO, USAJMO, IMO

An Official Publication of the **Mathematical Association of America**

EDITORIAL POLICY

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

A more detailed statement of author guidelines appears in this *Magazine*, Vol. 83, at pages 73–74, and is available at the *Magazine*'s website www.maa.org/pubs/mathmag.html. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, or published by another journal or publisher.

Please submit new manuscripts by email directly to the editor at mathmag@maa.org. A brief message containing contact information and with an attached PDF file is preferred. Word-processor and DVI files can also be considered. Alternatively, manuscripts may be mailed to Mathematics Magazine, 132 Bodine Rd., Berwyn, PA 19312-1027. If possible, please include an email address for further correspondence.

Cover image by Susan Stromquist

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

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

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

MAA Service Center, *maahq@maa.org*

All at the address:

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

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

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

Printed in the United States of America

Vol. 84, No. 4, October 2011

MATHEMATICS MAGAZINE

LETTER FROM THE EDITOR

The Olympiads We present the problems, solutions, and results from this summer's USAMO, USAJMO, and IMO. We thank the authors, who have had only a short time to prepare these features.

These competitions are a fine spectator sport, and if you have the time to try the problems for yourself, a participatory sport as well. It is nice to see the emergence of a few young superstars—you can read their names on pages 314, 315, 319—whom even the most rich and famous must envy.

But the competitions, from the first levels to the IMO, serve a more vital function. They help talented young mathematicians, especially those without good mathematical connections, to get the attention and the support that they need. There are many people for whom success in the contests has opened doors. I am one of them. Of course mathematics is much more than problem solving, and we need to have many ways for young people to demonstrate their talent. The contests are one way.

In This Issue Jacob Siehler describes a solitaire game played on a gameboard of 13 points. The points form a finite geometry, being organized into 13 "lines" of four points each. The game was invented by John Conway to explain certain simple groups. Now Siehler shows how actually to play the game, and how to win it.

Jonathan Hodge's article is about referendum elections. What if there are two or more issues on the ballot, but your preference on one issue depends on how the other issues are resolved? That is the starting point for adventures in combinatorics and modeling.

William E. Wood's title is "Squigonometry." To help us appreciate the ordinary sine and cosine, he puts them in a new light by developing what amounts to a second example of the same construction.

Frank Farris, past editor of this MAGAZINE, has kind words, set to music, for authors and referees.

In the notes we find another game: It is played between Cinderella and her stepmother, and it involves overflowing buckets. We also find two papers that both mention folding, but in different spaces.

One More Problem What if Cinderella and her stepmother (p. 278) play their game on Conway's gameboard? The stepmother distributes her gallon of water among 13 buckets, and Cinderella empties the four buckets from any "line" she chooses. How large must the buckets be for Cinderella to keep them from overflowing? If you find out, please let me know.

The Allendoerfer Awards We congratulate the winners of the 2011 Carl B. Allendoerfer Awards, for articles published in this MAGAZINE during 2010. They are Curtis Bennett, Blake Mellor, Patrick D. Shanahan, Gene Abrams, and Jessica K. Sklar.

Walter Stromquist, Editor

ARTICLES

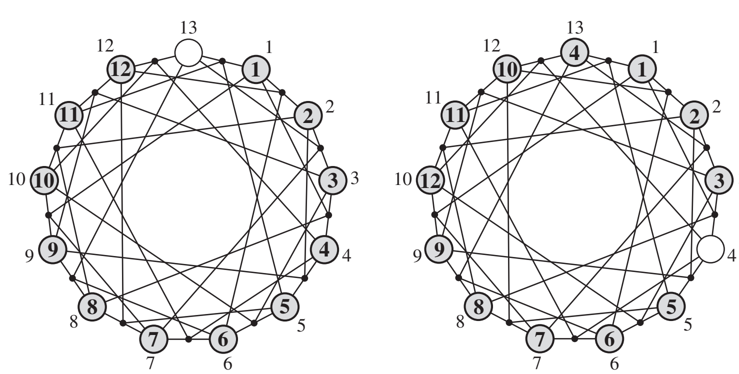
Depth and Symmetry in Conway's M_{13} Puzzle

JACOB A. SIEHLER
Washington & Lee University
Lexington, VA 24450
siehlerj@wlu.edu

Have you met M_{13}?

In a short 1997 article [**5**], John Conway describes a sliding-piece puzzle called M_{13}, which bears the same relation to the Mathieu group M_{12} as the famous 15-Puzzle bears to the alternating group A_{15} (the latter summed up neatly by Aaron Archer [**2**]). Much more information on the puzzle appeared in a subsequent, highly enjoyable article by Conway, Elkies, and Martin [**8**]. My first reaction on discovering the M_{13} articles was a somewhat unscholarly exclamation ("Play!"), so I implemented a clickable version of the puzzle on my computer [**16**]. A 15-puzzle veteran, I assumed I was equipped to solve M_{13} with minimal effort, but my endgame was vexed by the tricky moves. Should you find it similarly vexing to solve by hand, you will find a solution in the final section. However, I also found the game an excellent way to learn a bit more about the inner (and outer) workings of two much-renowned groups. The main results of this article will explain which algebraic symmetries of the abstract group M_{12} are compatible, in a certain sense, with the puzzle structure of M_{13} by connecting them to geometric symmetries of the playing surface.

The board for Conway's game is a "projective plane of order 3," which consists of thirteen points and thirteen lines. Each line contains four points, each point belongs to four lines, and any two lines intersect in a unique point. We won't worry about any more general definition of projective planes; you can read all about them in Stevenson's book [**17**] or many others. FIGURE 1 shows the particular 13-point plane that we will be using.



The starting state   One slide away from start

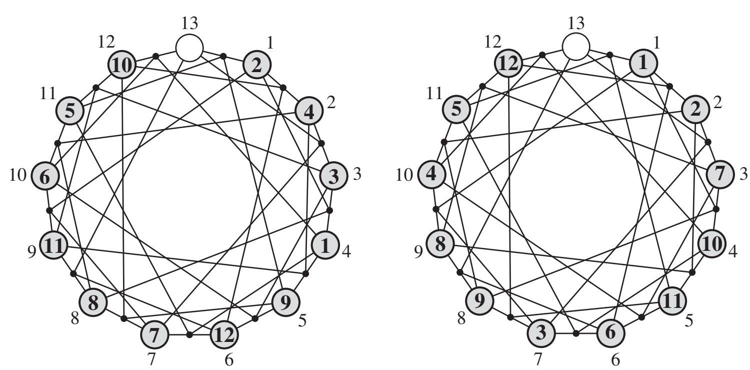
Figure 1 Numbered tiles on a projective plane of order 3

Math. Mag. **84** (2011) 243–256. doi:10.4169/math.mag.84.4.243.

The points of the plane are numbered 1–13, clockwise around the circle, and points 1–12 begin the game with like-numbered tiles on them. Point 13 begins with no tile upon it, and is thus deemed, for the time being, the "holy point." Lines of the plane are depicted by the dots around the circle, each dot connected to four points of the plane which constitute a line. With our numbering scheme, lines of the plane all have the form $\{i, i+1, i+5, i-2\}$, where i ranges from 1 to 13, and all the numbers are reduced mod 13.

There is only one type of legal move in the game: Choose any point of the plane with a tile on it, and move that tile to the current holy point (which incidentally makes your chosen point the new holy point). At the same time, swap the two tiles on the other points of the line which connects your point to the holy point. FIGURE 1 illustrates what happens if we move tile 4 to the holy point from the initial configuration. Pay attention to points 10 and 12, which complete the line containing 4 and 13.

Conceptually, it can be more convenient to think of the hole moving, and to describe a longer sequence of moves we'll simply list in succession the points the hole visits, in square brackets. For example, FIGURE 2 shows the results of playing two longer sequences, [1, 2, 4, 13] and [1, 2, 13, 6, 12, 13]. Verify these to check your understanding of the rules.



[1, 2, 4, 13] [1, 2, 13, 6, 12, 13]

Figure 2 Two move sequences and the resulting permutations of the tiles

Since both of these sequences return the hole to point 13, they can be viewed as permuting the twelve numbered tiles. In cycle notation, the first sequence induces the permutation (1 4 2)(5 11 9)(6 10 12), and the second induces (3 7)(4 10)(5 11)(8 9). Note how to read the cycles in game terms: whatever position you happen to start with, the first sequence will send the tile on point 1 to point 4; the tile on point 4 to point 2; and so on. Sequences such as these which return the hole to point 13 are called *closed* sequences, and the corresponding permutations can be composed to form a subgroup of S_{12}, which we'll call the *game group*, or G for short. If we think of the hole as just another tile numbered 13, to which special rules apply, then we can think of the game group as a subgroup of S_{13} (whose elements all leave the 13th tile fixed). It is properly contained in the set of all permutations in S_{13} that can be achieved by legal moves. This larger set is what Conway refers to as M_{13}. M_{13} is not a closed set under composition of permutations (the rules of the game place restrictions on which permutations can be legitimately composed), and it is not our primary interest here, although it was this larger set which originally motivated the game.

The point of all this as a *puzzle*, of course, is to start from a scrambled position such as the ones in FIGURE 2 and find a sequence of moves that returns the tiles to their initial positions. The slippery, indirect moves make this a challenge to solve intuitively, even though the puzzle has far fewer positions than other familiar permutation puzzles such as Rubik's Cube or the Fifteen Puzzle mentioned above.

One game, two simple groups

M_{13} and the 15-puzzle both generate interesting permutation groups as their pieces are scrambled. But compared to the rectangular grid of the 15-puzzle, the plane on which M_{13} tiles move has a rich group of symmetries in its own right. It is primarily the interplay between that group of geometric symmetries and the game group G (determined by the tile-sliding rules) which I want to explore. The next two theorems gather up the basic facts about these two groups.

THEOREM 1. *The game group is isomorphic to the Mathieu group M_{12}, a simple group of order* 95,040 *which acts sharply* 5*-transitively on a* 12*-element set.*

Just what is M_{12}? Part of the reason it still fascinates us more than a century after its discovery is that there are many equally good answers to that question, involving everything from twisting and sliding puzzles to card shuffling to codes. Joyner gives a nice view of M_{12}'s many faces [**12**], and there is still more in Conway and Sloane's weighty reference [**7**]. For the present purposes, it suffices to know that M_{12} is our game group, and that its deep and varied connections make it worth exploring from any accessible angle.

The isomorphism assertion connecting this game to other definitions of M_{12} is proven in the paper of Conway, Elkies, and Martin [**8**]. "Acts sharply 5-transitively" simply means the following: choose your favorite five tiles and your favorite five points for them to live on (an ordered choice). Then there is one and only one position in the game that puts your chosen five tiles on your chosen five points. For example, FIGURE 2 shows the one and only legal position with tiles 2, 4, 3, 1, and 9 on the first five points (in that order). This may sound like an unremarkable property for a group to have, but quite the opposite is true: The property of acting sharply 5-transitively on a 12-element set characterizes this group uniquely up to isomorphism. Dixon and Mortimer [**9**] give a nice summary of sharply k-transitive group actions, in addition to their easy-to-follow construction of M_{12} and the other Mathieu groups.

We can (and we will) understand some features of the game group better if we also pay attention to the symmetries of the playing surface—the projective plane. That gives us another set to consider inside S_{13}, namely, the permutations in S_{13} which preserve the plane's structure (entirely independent of whether they can be achieved by legal moves in the game). Preserving structure in this sense means that lines must be mapped to lines, and permutations that do this will be called *plane automorphisms*. For example, the picture of the plane we've used in FIGURE 1 makes at least one small family of plane automorphisms clear: we can simply rotate the figure one or more clicks clockwise, inducing a 13-cycle of the points which clearly sends each line to another line. Such permutations can be composed to give a group, which we'll call the *plane group*. It is known by various notations, including $PGL(3, 3)$, which is most suggestive of its role here. Much the way that an ordinary linear transformation can be described by its action on a basis, an element of our plane group can be described by its action on a suitable set of points in the plane: an *ordered oval* is an ordered set of 4 points in the plane, no 3 of which belong to a line.

THEOREM 2. *The plane group is a simple group of order* 5,616 *which acts sharply transitively on ordered ovals.*

This group belongs to an infinite family of finite simple groups, the projective special linear groups, and you can find an accessible proof that they are simple in the textbook of Beachy and Blair [**3**]. The transitivity assertion is proven in an elementary way in [**8**].

In both of the preceding theorems, the order of the group can be inferred from the transitivity assertion, as follows. In the case of M_{12}, any given selection of 5 tiles can be put in $12 \times 11 \times 10 \times 9 \times 8 = 95{,}040$ different positions (and elements of the group are in one-to-one correspondence with such ordered selections). In the case of the plane group, you need to count the number of ordered ovals: choose any two points (13×12), choose a third point off the line they determine (9 possibilities), and one more point off the three lines that have been determined so far (4 possibilities); all together, $13 \times 12 \times 9 \times 4 = 5{,}616$ ordered ovals, in one-to-one correspondence with elements of the plane group. The plane group also acts transitively on the points of the plane (this follows from transitivity on ovals, or more simply from the order 13 rotational symmetry we've already pointed out), so the stabilizer of any particular point—say, number 13—is a subgroup of order $5616/13 = 432 = 2^4 \cdot 3^3$.

Geometrically related sequences

As an example of interaction between the game and the plane group, let's look at some ways we can transform move sequences while preserving the cycle structure of their permutations.

An easy first example of such a transformation is the reversal of sequences: The inverse of $[a_1, a_2, \ldots, a_n, 13]$ is achieved by $[a_n, a_{n-1}, \ldots, a_1, 13]$, so these two sequences induce permutations of the same cycle type. The reversal would look more literal if I followed the convention of Conway and included a 13 at the beginning of the sequence to indicate the starting position of the hole, but I will still just refer to this as reversal.

Earlier I suggested that we might think of the hole as simply an invisible "tile number 13," to which special rules apply. We can easily imagine a variant game in which another numbered tile is the special one—call it an x-type game if tile x plays the "holy role." With this in mind, consider a move sequence $S = [a_1, a_2, \ldots, a_n]$ and a plane automorphism ϕ. We can apply ϕ term-by-term to S to obtain $\phi(S) = [\phi(a_1), \phi(a_2), \ldots, \phi(a_n)]$. If S is a closed sequence, inducing an element in the game group, then $\phi(S)$ is a closed sequence for a $\phi(13)$-type game, and would induce an element in the corresponding game group. This would have the same cycle type as S—the cycles of $\phi(S)$ are merely the cycles of S, "relabelled" by ϕ. We could say that two sequences related by a plane automorphism in this way are isomorphic sequences; this is an equivalence relation, and what we have just said as that isomorphic sequences induce the same type of permutation.

There is a weaker equivalence relation, slightly less obvious, which still gives the same implication for induced permutations. Consider a move sequence for the basic game, say,

$$S = [a_1, a_2, \ldots, a_n, 13]$$

If we rotate this sequence one step to the left to obtain

$$S_1 = [a_2, a_3, \ldots, a_n, 13, a_1],$$

we see that S_1 is a closed sequence for an a_1-type game. It will induce a permutation in that game of the same type as S. To see this, it helps to temporarily adopt an expanded notation for moves—let a parenthesized pair (xy) denote sliding the hole from x to y. Using this notation,

$$\begin{aligned} S_1 &= (13a_1)(a_1a_2)(a_2a_3)\cdots(a_n13) \\ &= (a_113)(13a_1)(a_1a_2)(a_2a_3)\cdots(a_n13)(13a_1) \\ &= (a_113)\ S\ (13a_1), \end{aligned}$$

and so S_1 is simply a conjugate of S, which will induce a permutation of the same cycle-type (on an a_n-type game) as S does on the ordinary game. This can be iterated; if S_i denotes the move sequence S rotated i terms to the left, then S_i is a closed sequence for an a_i-type game which induces a permutation of the same cycle type as S does on the ordinary game.

PROPOSITION 1. *If S and T are closed sequences (for the standard game) and T can be obtained from S by (1) applying any plane automorphism to the terms of S, (2) reversing and/or rotating the resulting move sequence any number of positions left or right, then T and S induce permutations of the same cycle type.*

This gives each move sequence for the ordinary game a large class of sequences of the same length which behave "the same," at least as far as cycle structure. This can be helpful in reducing the amount of computation needed to study move sequences of moderate length.

Plane automorphisms within the game group

Once I learned to play the game well and unscramble the tiles reliably, I began looking for other amusements, and the question of realizing plane automorphisms via legal moves in the game was a challenge that occurred naturally—somewhat like the pastime of constructing pretty patterns on Rubik's cube. The prettiest patterns on the cube, generally, are the ones that capitalize on its symmetry in some way.

Of course, most permutations in M_{13}'s game group are not plane automorphisms. The first example in FIGURE 2 is typical: the line connecting points 1 and 2 in the plane now contains tiles {2, 4, 10, 12} which do *not* constitute a line of the plane. Of course, that must be true simply by cardinality (the game group is far larger than even the entire plane group), but experimenting with short move sequences might make you wonder whether there are *any* legal moves which send the tiles of each line to the points of another line. Consider the second example in FIGURE 2, however. If you find it tedious to inspect all thirteen lines to verify that this move does indeed result in a plane automorphism, a suitable drawing of the plane (FIGURE 3) can reveal it "at a glance" as a reflection of the plane across the line {1, 2, 6, 12}. In fact, any sequence of the form $[a, b, 13, c, d, 13]$ where $\{a, b, c, d\}$ form a line *not* incident with point 13 will induce a plane automorphism with a similar reflection picture. In FIGURE 3, we've taken the liberty of doubling points on the line {2, 3, 7, 13} in order to put as much symmetry on display as possible. If you like, you can imagine stitching the boundary together so that antipodal points are identified, eliminating the doubling—and incidentally embedding our diagram in the real projective plane $\mathbb{R}P^2$—but you'll have to imagine it on very stretchy material. The model of the plane used in FIGURE 3 can be found in many references on projective planes. Polster's book [**14**] gives it and several other useful pictures that we'll be using here.

Once you've found a few automorphisms in the game group, it's hard to resist wondering just how many you can find. I'd like to answer that question gradually, starting with a very distinguished set. The paper of Conway, Elkies, and Martin [**8**] identifies

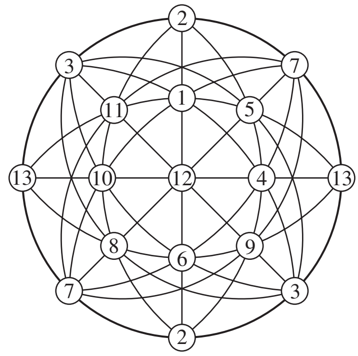


Figure 3 Projective plane exhibiting reflective symmetries

eight positions in the game group which require the maximum number of moves—nine moves, as they show there—to achieve from the start position (or equivalently, to return to the start position). Together with the identity, those eight form an abelian group of order 9, in which every nonidentity element has order 3. We denote this group T; it is generated by

$$\theta = [9, 1, 13, 2, 4, 3, 7, 10, 13] = (1\ 5\ 11)(6\ 9\ 8)(4\ 10\ 12)$$

and

$$\omega = [10, 1, 11, 12, 5, 13, 7, 9, 13] = (2\ 3\ 7)(4\ 10\ 12)(6\ 8\ 9).$$

Once it occurs to you to check, you can verify that θ and ω are both elements of the plane group, but in the case of θ, I hope the picture of the plane in FIGURE 4 is more enlightening than a line-by-line audit. Several points (numbers 2, 3, and 7) may appear to be missing from FIGURE 4, but they are tucked away behind 13 on an axis perpendicular to the printed page. Also, the circular "lines" each seem to be missing a point, but each of them contains one of the points 2, 3, or 7 on that central axis. This picture is uncomfortable in two dimensions; it would prefer to be seen as a torus, with the fixed points on the central axis and other lines of the plane as "lines" of slope 0, 1, and -1 wrapping around it. But either way you picture it, there is a rotational symmetry of order 3 which is precisely θ, and you can provide a similar picture to show that ω can also be considered as an order 3 rotational symmetry of the plane. It follows that all the elements of the "maximum depth group" T belong to the plane

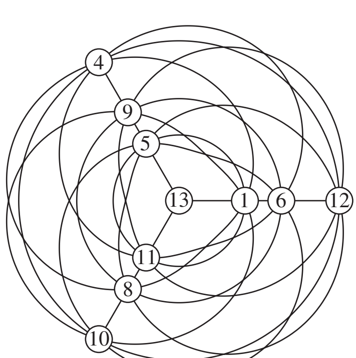


Figure 4 Rotational symmetry of order 3 about a fixed line

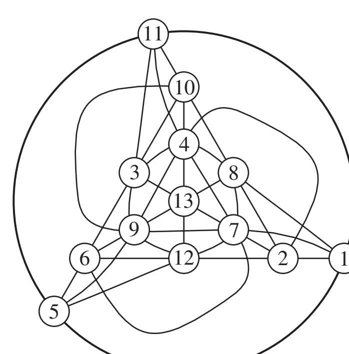


Figure 5 Rotational symmetry of order 3 about a single fixed point

group as well, and they all have similar geometry: one line through point 13 is fixed (pointwise) as the axis of rotation, while the other three points in each of the remaining lines through point 13 move in a 3-cycle. Now, consider one more move sequence and its associated permutation,

$$\zeta = [4, 9, 11, 6, 8, 10, 5, 13] = (1\ 11\ 5)(2\ 10\ 6)(3\ 12\ 8)(4\ 9\ 7),$$

and yet another picture of our plane (FIGURE 5) to show that ζ is also a plane automorphism.

This ζ normalizes the group T generated by θ and ω—you can check this directly, or see Proposition 3 below. All together, then, they generate a group of order 27 which belongs to the intersection of the game group and the plane group. Since $432 = 2^4 \cdot 3^3$, we have found a Sylow 3-subgroup in that intersection.

In constructing a Sylow 2-subgroup in the intersection, we will again rely on figures. FIGURE 6 shows one more, partially completed model of the projective plane. Rotating the hook-shaped "line" through multiples of $\pi/4$ supplies the remaining lines of the projective plane, which I'll omit for clarity.

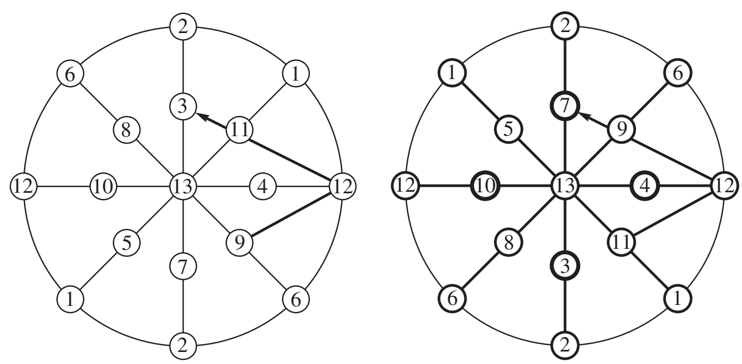


Figure 6 Partial diagram of the plane. Reflection doesn't carry lines to lines...

The completed model obviously has a rotational symmetry of order 8; call it ρ. In cycle form, $\rho = (1\ 2\ 6\ 12)(3\ 8\ 10\ 5\ 7\ 9\ 4\ 11)$, and this can be achieved in the game group as $\rho = [7, 6, 8, 12, 6, 3, 4, 13]$. The figure may appear to have reflective symmetry across the four lines that appear as diameters of the circle, but this is not the case. Simply reflecting points across, say, the horizontal diameter as shown, we find that the line {9, 12, 11, 13} has *not* been carried to another line of the plane. The lines in the model have a counterclockwise orientation, if you follow them from short end to long end as indicated by the arrowheads. Reflection reverses that orientation. However, we can fix this with just a small adjustment: give the four emphasized points in FIGURE 6 an additional 180° twist, and all the lines have been restored to their proper orientation in FIGURE 7. I have left just one of the hook-shaped lines in the diagram; you can pencil in its rotations to check that they are actually lines of the plane.

Let's call this reflection-with-a-twist $\tau = (1\ 6)(5\ 8)(9\ 11)(4\ 10)$. It belongs to the game group as, for example, $\tau = [12, 7, 3, 12, 2, 13]$. You can compute that $\tau\rho\tau^{-1} = \rho^3$ (use the picture!), so τ and ρ generate a group of order 16 which is a Sylow subgroup in the intersection of the game group and the plane group. If τ were literally a reflective symmetry of this model, we would have the familiar dihedral group of order 16; as it is, our reflection-with-a-twist generates what is known as the *quasidihedral group* or *semidihedral group* of order 16. Usage varies, but the latter term is used by

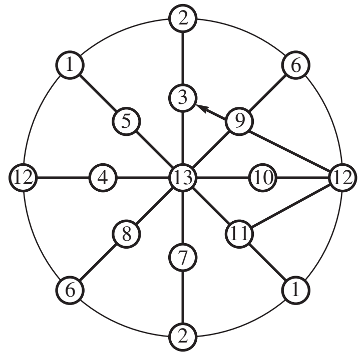


Figure 7 . . . but that can be corrected with an additional twist.

Wild [**18**] in his round-up of groups of order 16. The projective plane diagrams above are the simplest geometric representation I know for this group. And with plane automorphism groups of order 3^3 and 2^4 inside the game group, we have proven the following:

PROPOSITION 2. *The game group contains all* 432 *plane automorphisms which leave point* 13 *fixed.*

For the remainder of the paper, let P denote this 432-element subgroup of the game group. P can be presented with just two generators; ζ and τ will suffice, in fact, and with access to computer algebra systems it is easy to find such pairs of generators—but it is also satisfying to see automorphisms which generate the Sylow subgroups, concretely and geometrically. Both the game group of order 95040 and the plane group of order 5616 can also be presented with two generators. See the short articles of Leech [**13**] and Bisshopp [**4**] for specific generators, if you like.

A few terms and definitions

These plane automorphisms within the game group play a special role in the following sections, where we will study group automorphisms of M_{12} and how they relate to the extra structure provided by the rules of Conway's game. A quick review of some standard definitions and facts may be useful before we go on. Since we've committed the letter G to denoting our game group, let H stand for any finite group.

The set $Z(H) = \{x \in H | xh = hx, \forall h \in H\}$ is known as the *center* of H. It's routine to check that $Z(H)$ is a normal subgroup of H, so, in particular, the center of a nonabelian simple group like G is trivial.

The automorphisms of H themselves form a group under composition, which is denoted $\mathrm{Aut}(H)$. An *inner automorphism* of H is an automorphism of the form $x \mapsto hxh^{-1}$ ("conjugation by h") for some fixed $h \in H$. These form a subgroup of $\mathrm{Aut}(H)$, denoted $\mathrm{Inn}(H)$. There is a surjective homomorphism from H to $\mathrm{Inn}(H)$ (namely, mapping each $h \in H$ to the "conjugation by h" automorphism), the kernel of which is none other than $Z(H)$. Thus, if $Z(H)$ happens to be trivial (as in the case of G), then this mapping is injective; each element of H represents a distinct element of $\mathrm{Inn}(H)$.

Elements of $\mathrm{Aut}(H)$ which are not inner automorphisms are called *outer automorphisms*. A final routine fact is that $\mathrm{Inn}(H)$ is a normal subgroup of $\mathrm{Aut}(H)$, and the quotient group $\mathrm{Aut}(H)/\mathrm{Inn}(H)$ is denoted $\mathrm{Out}(H)$, the *outer automorphism group* of H. Confusingly, the elements of $\mathrm{Out}(H)$ are not outer automorphisms; they are equivalence classes of automorphisms. Two elements α and β in $\mathrm{Aut}(H)$ represent the same class in Out(H) if and only if α can be expressed as the composition of β with some inner automorphism of H.

With definitions in hand, it is now time to identify, precisely, the "extra structure" that I have been referring to.

Depth and automorphisms of the game group

The *depth* of a position in the game group is the minimum number of moves required to solve that position (or, equivalently, to achieve it from the initial state). Depth is not a purely group-theoretic notion. It comes, essentially, from the choice of generators that we are allowed by the legal moves of the game—and that means that there's no reason to expect automorphisms of the game group to preserve depth. Elements of the subgroup P, however, do give us examples:

PROPOSITION 3. *Conjugation by a plane automorphism in the game group preserves depth: That is, if $\phi \in P$ and $\alpha \in G$, then α and $\phi\alpha\phi^{-1}$ have the same depth.*

Proof. Let $S = [a_1, a_2, \ldots, a_n, 13]$ be a sequence of moves that induces α, and consider $\phi(S) = [\phi(a_1), \phi(a_2), \ldots, \phi(a_n), 13]$. It's easy to see how $\phi(S)$ must act on the tiles: if S carries the tile on point x to point y, then $\phi(S)$ carries the tile on point $\phi(x)$ to point $\phi(y)$. But that is precisely how $\phi\alpha\phi^{-1}$ acts, so $\phi(S)$ is a sequence that achieves $\phi\alpha\phi^{-1}$. Thus for every sequence that achieves α there is a sequence of the same length that achieves $\phi\alpha\phi^{-1}$, and vice versa of course, using ϕ^{-1} in place of ϕ. ■

In the present case, conjugation by elements of P provides 432 inner automorphisms of G which preserve depth (432 *distinct* automorphisms, in fact; you can deduce this from the fact that G is simple). And that's all there are:

PROPOSITION 4. *G has precisely* 432 *inner automorphisms which preserve depth.*

Proof. Since the previous proposition shows there at least 432, we only have to show that there are no more than 432. Consider the element

$$\theta = [9, 1, 13, 2, 4, 3, 7, 10, 13] = (1\ 5\ 11)(6\ 9\ 8)(4\ 10\ 12)$$

We have already mentioned that all the (nonindentity) elements of the maximum-depth group T have the same cycle structure as θ—the fixed points constitute a line through point 13, and each of the remaining lines through 13 has its three remaining points permuted in a three-cycle. Thus, if conjugation by α carries θ to another element of T, α must carry the set $\{1, 5, 11\}$ to a set of three points which are collinear with 13, and there are 12×2 ways to do this. Once the image of that set is determined, there are only six choices for $\alpha(6)$ and then three choices for $\alpha(4)$. And at this point, if α belongs to G, it is uniquely determined, since G is sharply 5-transitive. Thus there are at most $12 \times 2 \times 6 \times 3 = 432$ such elements α in G. ■

Outer automorphisms

Our game group has more than inner automorphisms, but not much more; it is a fact that the outer automorphism group of M_{12} has order two [**6**]. That means any two outer automorphisms of M_{12} represent the same class in Out(M_{12}), and differ from one another only by composition with some inner automorphism. Conway et al. [**8**] show how one such outer automorphism can be discovered using a "dualized" version of the M_{13} game in which one slides lines as well as points. We will not duplicate that exposition here, but simply rely on their work to produce an explicit representative of the nontrivial class of outer automorphisms. To describe our automorphism, we'll need a set of generators for the game group. The following two (chosen fairly arbitrarily) suffice.

$$x = [4, 11, 3, 4, 13] = (1\ 8)(3\ 11)(6\ 7)(9\ 10)$$

and

$$y = [8, 7, 11, 3, 4, 12, 1, 13] = (1\ 8\ 10)(2\ 6\ 4)(3\ 5\ 9)(7\ 12\ 11)$$

We won't verify directly that they are generators, but they satisfy the relations for the presentation given in the online *Atlas of Finite Group Representations* [**1**]. By following the dualization construction (making the necessary adjustments for the different labeling of points in our projective plane) we can produce a specific outer automorphism of the game group, namely F, which acts on the generators above by

$$F(x) = (1\ 10)(2\ 8)(3\ 6)(5\ 11)$$

and

$$F(y) = (1\ 12\ 6)(4\ 9\ 5)(8\ 11\ 10)$$

You can check that these satisfy the same relations as x and y, so we do have an automorphism (it's easily checked once it's presented, but finding it would be difficult without a clever construction such as dualizing). Is it depth-preserving? Recall our maximum-depth element $\theta = (1\ 5\ 11)(6\ 9\ 8)(4\ 10\ 12)$ from the section on plane automorphisms. We can express θ in terms of our generators, $\theta = (xy)^3y(xy)^{-3}y^{-1}$, from which we can compute $F(\theta) = (1\ 12\ 6)(4\ 9\ 5)(8\ 11\ 10)$, which is not one of the eight elements in subgroup T, so F is not depth-preserving. But note that a similar calculation with generators shows that $F(\omega) = \theta$, so F does send ω to another element at depth 9.

It is still plausible that some other outer automorphism of G might be depth-preserving. A little computer assistance helps to rule this out; the computer can easily give us a list of the conjugacy classes and their sizes in the game group. The list of 15 classes in TABLE 1 shows each class's cycle type in addition to its size. The same information is available in the *Atlas* [**6**].

TABLE 1: Conjugacy classes in M_{12} (arbitrarily numbered)

Class	Cycle type	Size	Class	Cycle type	Size
1	Identity	1	9	3^4	2640
2	2^4	495	10	4^2	2970
3	2^6	396	11	$4 \cdot 8$	11880
4	$2^2 \cdot 4^2$	2970	12	5^2	9504
5	$2 \cdot 3 \cdot 6$	15840	13	6^2	7920
6	$2 \cdot 8$	11880	14	11	8640
7	$2 \cdot 10$	9504	15	11	8640
8	3^3	1760			

Now, consider an element from the 4^2 class, such as $z = [2, 10, 3, 7, 8, 13] = (5\ 9\ 6\ 12)(7\ 10\ 8\ 11)$. By expressing z in terms of x and y (a little tedious) we can compute that $F(z) = (1\ 12)(7\ 10)(2\ 3\ 8\ 11)(4\ 5\ 9\ 6)$, in the $2^2 \cdot 4^2$ class, so F exchanges these two classes, and every outer automorphism of G must do the same. Those classes, as it turns out, both contain elements at depth 6 and 7 only. But they differ in their distribution. See TABLE 2, which was computed with Mathematica. Thus, every outer automorphism of G moves elements from depth 6 to depth 7 and vice versa. And with that, we have proven:

TABLE 2: Distribution by depth for two conjugacy classes

Class	Cycle type	Depth 6	Depth 7
4	$2^2 \cdot 4^2$	972	1998
10	4^2	1188	1782

PROPOSITION 5. *The game group G has precisely* 432 *automorphisms which preserve the depth of every element in the group, namely, conjugation by the* 432 *elements of the plane automorphism subgroup P.*

We can now deduce one further result which identifies the subgroup T as the essential "test case" for depth-preserving automorphisms. The proof of Proposition 4 actually shows that an inner automorphism which preserves the subgroup T preserves the depth of every element in G. And as we have remarked, any outer automorphism of M_{12} is simply the composition of our particular F with some inner automorphism. Consider F followed by conjugation with some element g. $F(\omega) = \theta \in T$, so if $gF(\omega)g^{-1}$ also belongs to T, then we know that g is a plane automorphism, and $gF(x)g^{-1}$ has the same depth as $F(x)$ for every $x \in G$. In particular, we computed that $F(\theta) \notin T$, so $gF(\theta)g^{-1} \notin T$. Thus, every outer automorphism sends some element of T, either ω or θ, to an element of strictly lower depth, and we have proven the following tidy (and surprising) result.

PROPOSITION 6. *An automorphism of the game group G preserves the depth of all elements if and only if it preserves depth for the eight maximal-depth elements in the group.*

It would be interesting to see if a similar result holds for other permutation puzzles, such as the Skewb or the Pyraminx, which have a small number of positions at the maximum depth. These puzzles are described in Hofstadter's popular article [**11**] and Joyner's more recent book [**12**], but the best source for information about the distribution by depth in these puzzles is online [**15**].

Solving M_{13} by hand

In one sense, the "optimal" algorithm for solving the puzzle is what puzzlers sometimes refer to as God's Algorithm (this term was at least popularized, if not originated, by Hofstadter [**10**])—since there are a finite number of positions, simply memorize the shortest solution for each one! This is not practical for human puzzlers, however, and the algorithm given by Conway et al. [**8**] is also not practical if we don't wish to rely on computerized assistance. What follows is an admittedly idiosyncratic set of moves that can be memorized easily and will suffice to unscramble the puzzle from any scrambled state (provided that the scrambled state was obtained by legal moves in the first place). In order to guarantee visible progress as we proceed through the puzzle we will position tiles one at a time, and rely on the magic of sharp transitivity to ensure that the puzzle is solved precisely when we slide the fifth tile into its proper position. Of course, at each step we want to preserve the tiles we have already positioned correctly, and that means we work in progressively smaller subgroups—the stabilizers of our growing set of correctly positioned tiles.

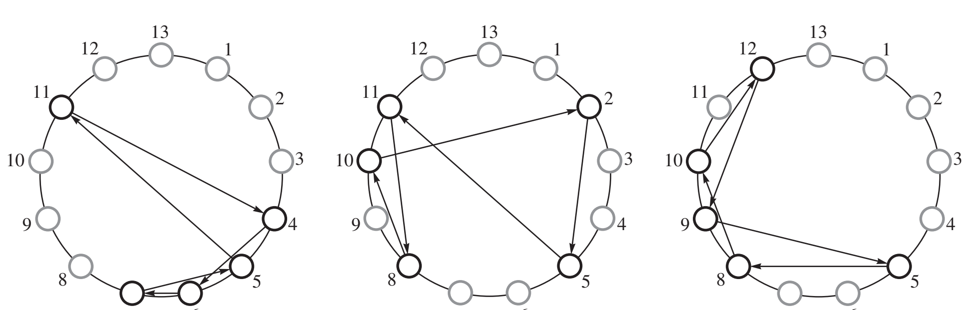
Stage 1: Position tile 1 That is, position tile 1 on point 1, leaving point 13 unoccupied. This can be done intuitively from any position with a closed sequence of just 3 moves, and is left as an exercise to ensure familiarity with the rules of the game. Although this is simple, you'll find you need to vary your strategy depending on whether the point which holds tile 1 is colinear with points 1 and 13, or not (with the colinear case requiring a bit more indirection).

Stage 2: Position tile 3 From this point on, we will only use move sequences that fix the tile on point 1. That restricts us to a subgroup of order 11!/7!—which is, incidentally, the group known as M_{11}. Since 11 is a prime divisor of the order, there must be an element of order 11 in the group by the well known theorem of Cauchy. Such an element could only act as an 11-cycle on tiles 2–12. Thus the puzzler who has a rabid preference for minimizing the number of cases to consider need only find a short sequence of moves which induces such an 11-cycle, and repeat that sequence until tile 3 is returned home. For example, either of the sequences

$$[4, 5, 6, 7, 8, 13] \quad \text{or} \quad [2, 4, 6, 8, 10, 13]$$

is very easy to remember and induces an 11-cycle leaving the tile on point 1 fixed. Simply repeat until tile 3 has returned home. Realistically, though, this stage requires only 3 or 4 moves, and can be carried out (after a little practice) without relying on memorized sequences.

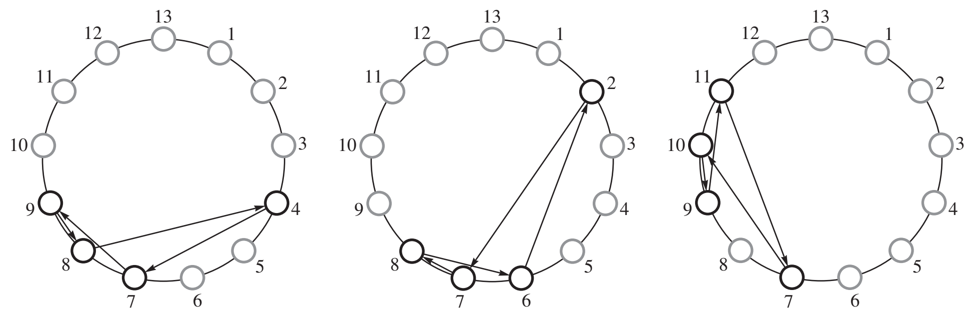
Stage 3: Position tile 5 One of the sequences in FIGURE 8 can be applied (perhaps repeatedly) to return tile 5 home while preserving the tiles on points 1 and 3. Each of them actually induces a *double* 5-cycle, but the illustration shows only the cycle which moves tiles to and from point 5. Simply choose a cycle which contains tile 5 and repeat the sequence until it returns home.



$\alpha = [4, 5, 7, 6, 13]$ $\quad$ $\beta = [8, 5, 2, 10, 13]$ $\quad$ $\gamma = [4, 2, 11, 6, 13]$

Figure 8 Move sequences for Stage 3.

Stage 4: Position tile 7 Very similar to the previous stage: the three sequences in FIGURE 9 almost suffice to return tile 7 home. Again, each of them induces a *double* 4-cycle, but we've only shown the relevant part in the figure. The sequences are quite easy to remember, all having the form $[1, x, 5, 11, y, 13]$.



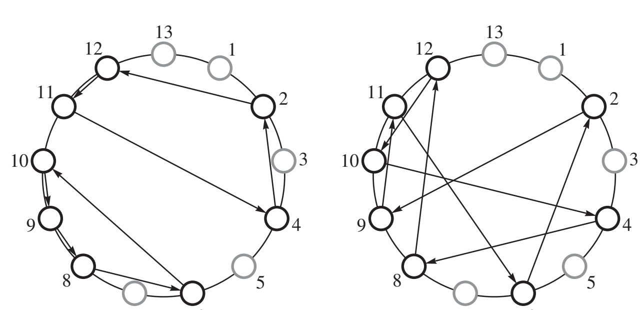
$\rho = [1, 12, 5, 11, 6, 13]$ $\sigma = [1, 12, 5, 11, 4, 13]$ $\tau = [1, 7, 5, 11, 10, 13]$

Figure 9 Move sequences for Stage 4.

Three 4-cycles don't quite suffice to cover all the possibilities. If tile 7 happens to be on point 12 at this stage, you can execute [8, 10, 13] to move it safely to point 4, and then apply ρ. (Since the stabilizer of $\{1, 3, 5\}$ has order 72, we can't use 5-cycles at this stage. There are elements of order 6 in the stabilizer, and you might hope to find a pair of 6-cycles which would cover all the possibilities at this stage, but a quick search on the computer shows this is not possible.)

Stage 5: Position one more tile At this stage we are trying to preserve the position of tiles 1, 3, 5, and 7, which means we are working in a subgroup of only 8 elements! The two sequences in FIGURE 10 suffice to generate the group (and it's easy to check that the group they generate is isomorphic to the familiar quaternion group of 8 elements). If some power of i or j will suffice to return any tile home, great! Otherwise execute i, and then either j or j^{-1} will complete the solution. If you should find that you have five tiles on the correct points but the puzzle still isn't solved, then it never will be solved by further legal moves; it's in a state that lies outside the game group. As a little optimization, note that i^2 and j^2 induce the same permutation of order 2, which can be achieved more efficiently by the sequence [8, 10, 13].

Working down through stabilizers in this way is a simple-minded strategy and is generally very far from providing the shortest possible solution. However, I do hope



$i = [6, 3, 9, 5, 10, 13]$ $j = [4, 1, 12, 3, 8, 13]$

Figure 10 Move sequences for Stage 5 you will have the opportunity to try it with a playable model of the puzzle. I have referred to the "magic" of sharp 5-transitivity, and the feeling that tiles are conspiring to come into place just as I make my last move is still eerie, even though I have seen, on paper, the proof that they must do so. M is for *Mathieu*, naturally, but a little time playing with the M_{13} puzzle can remind us to keep words like *marvelous* and *magical* in mind as well.

REFERENCES

1. Rachel Abbott et al., Atlas of Finite Group Representations, http://brauer.maths.qmul.ac.uk/Atlas/v3/
2. Aaron F. Archer, A modern treatment of the 15 puzzle, *Amer. Math. Monthly* **106** (1999) 793–799. http://dx.doi.org/10.2307/2589612
3. John A. Beachy and William D. Blair, *Abstract Algebra*, 3rd ed. chapter 7, 357–362. Waveland Press, 2006.
4. K. E. Bisshopp, Abstract defining relations for the simple group of order 5616, *Bull. Amer. Math. Soc.*, **37** (1931) 91–100. http://dx.doi.org/10.1090/S0002-9904-1931-05112-0
5. J. H. Conway, M_{13}. In *Surveys in Combinatorics, 1997*, 1–11. Cambridge University Press, 1997.
6. J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of Finite Groups*, Oxford University Press, 1985.
7. J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, chapter 11, pages 299–330. Springer-Verlag, 1988.
8. John H. Conway, Noam D. Elkies, and Jeremy L. Martin, The Mathieu group M_{12} and its pseudogroup extension M_{13}, *Experimental Mathematics* **15** (2006) 223–236. http://dx.doi.org/10.1080/10586458.2006.10128958
9. John D. Dixon and Brian Mortimer, *Permutation Groups*, Springer, 1996.
10. Douglas R. Hofstadter, Magic Cubology. In *Metamagical Themas*, Basic Books, 1985, 301–328.
11. ———, On Crossing the Rubicon. In *Metamagical Themas*, Basic Books, 1985, 329–363.
12. David Joyner, *Adventures in Group Theory*, 2nd ed., The Johns Hopkins University Press, 2008.
13. John Leech, A presentation of the Mathieu group M_{12}, *Canadian Mathematical Bulletin* **12** (1969) 41–43. http://dx.doi.org/10.4153/CMB-1969-005-8
14. Burkhard Polster, *A Geometrical Picture Book*, Springer, 1998.
15. Jaap Scherphuis, *Jaap's Puzzle Page*, http://www.jaapsch.net/puzzles/
16. Jacob A. Siehler, Conway's M(13) Puzzle. http://demonstrations.wolfram.com/ConwaysM13Puzzle/
17. Frederick W. Stevenson, *Projective Planes*, W. H. Freeman and Company, 1972.
18. Marcel Wild, The groups of order sixteen made easy, *Amer. Math. Monthly* **112** (2005) 20–31. http://dx.doi.org/10.2307/30037381

Summary We analyze a sliding tile puzzle (due to J. H. Conway) which gives a presentation of the Mathieu group M_{12}, one of the sporadic simple groups. We use the symmetries of a finite projective plane to classify automorphisms of M_{12} which preserve the depth of puzzle positions, and give an algorithm for solving the puzzle by hand.

JACOB SIEHLER earned his Ph.D. at Virginia Tech and teaches at Washington & Lee University, where his office shelves probably hold more textbooks than toys (but it's a close thing). He sometimes claims to be "studying fiber bundles" when he actually means taking an afternoon to play with his dogs.

Squigonometry

WILLIAM E. WOOD
University of Northern Iowa
Cedar Falls, IA 50614
bill.wood@uni.edu

It is easy to take the circle for granted. In this paper, we look to enhance our appreciation of the circle by developing an analog of trigonometry—a subject built upon analysis of the circle—for something that is *not quite* a circle. Our primary model is the unit *squircle,* the superellipse defined as the set of points (x, y) in the plane satisfying $x^4 + y^4 = 1$, depicted in FIGURE 1. It is a closed curve about the origin, but while any line through the origin is a line of symmetry of the circle, there are only four lines of symmetry for the squircle. Many familiar notions from trigonometry have natural analogs and we will see some interesting behaviors and results, but we will also see where the lower degree of symmetry inconveniences our new theory of *squigonometry*. We only scratch the surface here, offering many opportunities for the reader to extend the theory into studies of elliptic integrals, non-euclidean geometry, number theory, and complex analysis.

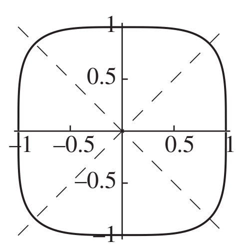


Figure 1 The unit squircle

Trigonometric functions and differential equations

We begin with the classical parameterization of the circle as the set of points of the form $(\cos t, \sin t)$ for $0 \le t \le 2\pi$. The cosine and sine functions report the x and y coordinates of the circle $x^2 + y^2 = 1$. There is nothing to stop us from doing the same thing for a squircle. We will define the functions as solutions to coupled initial value problems analogous to those that define the trigonometric functions. This approach is inspired by methods discussed in [**1**] and [**2**] of using IVP's to develop transcendental functions in a first-year calculus course.

Recall from calculus that $\frac{d}{dt}\cos t = -\sin t$ and $\frac{d}{dt}\sin t = \cos t$. We view these as defining properties for this pair of functions. When we combine these relationships with initial conditions, we can say that cosine and sine are the functions satisfying

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

where x corresponds to cosine and y to sine. This is an example of a *coupled initial value problem*, and it turns out that problems like this always have unique solutions.

Math. Mag. **84** (2011) 257–265. doi:10.4169/math.mag.84.4.257. © Mathematical Association of America

Therefore, we may *define* cosine and sine to be the unique solution to (1). We could then use Euler's method to approximate values, and the graph of the phase plane would trace out the unit circle. Further, the equations (1) are enough to derive all of the familiar properties of these functions.

For example, consider the function $f(t) = u(t)^2 + v(t)^2$, whose derivative is given by $f'(t) = 2u(t)u'(t) + 2v(t)v'(t)$. If $u(t) = \cos(t)$ and $v(t) = \sin(t)$ are defined by (1), then in this case $f'(t) = 0$ and $f(0) = 1$. A function whose derivative is always zero must be constant, so $f(t) = 1$ for all t, or $\cos^2 t + \sin^2 t = 1$. This is an example of a general method for proving identities with functions defined as solutions to (coupled) initial value problems.

As another example, we show how to use these techniques to prove the symmetries of cosine and sine. Let $\alpha(t) = \cos(-t)$ and $\beta(t) = -\sin(-t)$. Then $\alpha(0) = 1$, $\beta(0) = 0$, $\alpha'(t) = \sin(-t) = -\beta(t)$, and $\beta'(t) = \cos(-t) = \alpha(t)$. In other words, α and β are solutions to the CIVP (1) that defines sine and cosine. But the solution of that CIVP is unique, so it must be that $\cos(-t) = \alpha(t) = \cos(t)$ and $-\sin(-t) = \beta(t) = \sin(t)$.

We also know the tangent function as $\tan t = \frac{\sin t}{\cos t}$, and its derivative $\frac{d}{dt}\tan t = \frac{d}{dt}\frac{\sin t}{\cos t} = \frac{1}{\cos^2 t} = 1 + \tan^2 t$. We could thus redefine tangent as the solution to the initial value problem $\frac{dx}{dt} = 1 + x^2$, $x(0) = 0$, a familiar separable differential equation.

We now discuss this existence/uniqueness result in more generality. The theorems for the uncoupled and coupled cases are as follows:

THEOREM 1. *Let x_0, t_0 be real numbers and let $F(x, t)$ be continuous with continuous partial derivatives. Then there is a unique function $x(t)$ defined on an interval containing t_0 satisfying*

$$\begin{cases} x'(t) = F(x(t), t) \\ x(t_0) = x_0. \end{cases}$$

THEOREM 2. *Let x_0, y_0, and t_0 be real numbers and let $F(x, y, t)$ and $G(x, y, t)$ be continuous with continuous partial derivatives. Then there are unique functions $x(t)$ and $y(t)$ defined on an interval containing t_0 satisfying*

$$\begin{cases} x'(t) = F(x(t), y(t), t) \\ y'(t) = G(x(t), y(t), t) \\ x(t_0) = x_0 \\ y(t_0) = y_0. \end{cases}$$

It is worth noting that both theorems are really just special cases of a more general theorem which says that IVP's defined on vector-valued functions have unique solutions. Theorems 1 and 2 are just the one- and two-dimensional cases, respectively.

Parameterizing the squircle

To develop our theory of squigonometry, we must define functions that do for squircles what cosine and sine do for circles. If we are to use a coupled initial value problem to define our functions, the solutions $u(t)$ and $v(t)$ must make the function

$$g(t) = u(t)^4 + v(t)^4 \qquad (2)$$

constant. We design the following coupled initial value problem with that property in mind:

$$\begin{cases} x'(t) = -y(t)^3 \\ y'(t) = x(t)^3 \\ x(0) = 1 \\ y(0) = 0 \end{cases} \tag{3}$$

Thus, if $u(t)$ and $v(t)$ are the solution functions to (3), then $g'(t) = 4u(t)^3u'(t) + 4v(t)^3v'(t) = -4u(t)^3v(t)^3 + 4v(t)^3u(t)^3 = 0$, so $g(t)$ must be constant. Since $g(0) = 1$, it follows that $g(t)$ is identically one, as desired. We denote the unique pair of functions satisfying (3) by $\text{cq}(t) = x(t)$ and $\text{sq}(t) = y(t)$, the *cosquine* and *squine* functions, respectively. Again, note that Theorem 2 is what allows us to use the CIVP as a device to define these functions.

EXERCISE 1. *Prove that the cosquine and squine functions are even and odd, respectively.*

EXERCISE 2. *Use a computer algebra system to find the Maclaurin series for cosquine and squine.*

We also define the *tanquent* function $\text{tq}(t) = \frac{\text{sq}(t)}{\text{cq}(t)}$. Then

$$\frac{d}{dt}\text{tq}(t) = \frac{\text{sq}(t)^4 + \text{cq}(t)^4}{\text{cq}(t)^2} = \frac{1}{\text{cq}(t)^2} = \sqrt{\text{tq}(t)^4 + 1},$$

the last equality following from the easily verifiable identity $\text{tq}(t)^4 + 1 = \frac{1}{\text{cq}(t)^4}$. We see from their graphs in FIGURE 2 that the squigonemetric functions are "flatter" versions of their trigonometric analogs in much the same way as the squircle is a "flattened" circle.

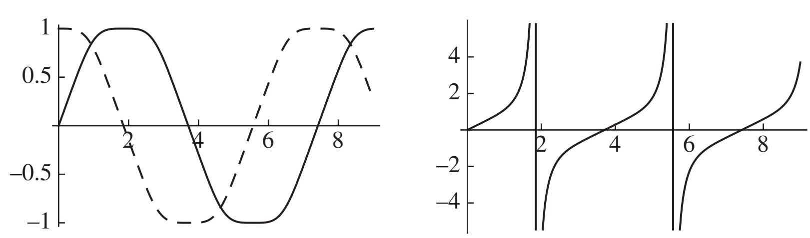


Figure 2 Plots of the squine (left, solid), cosquine (left, dashed), and tanquent (right)

EXERCISE 3. *By letting* $z = \frac{y}{x}$, $w = \frac{1}{x}$ *in* (1) *we obtain the following defining CIVP for the tangent and secant functions:*

$$\begin{cases} z'(t) = w^2 \\ w'(t) = z \\ z(0) = 0 \\ w(0) = 1. \end{cases}$$

Develop a CIVP for the corresponding squigonometric functions. Do the same for cotangent and cosecant and their analogs.

The differences between the trigonometric functions and their squigonometric counterparts may be illustrated by analyzing the radius and angle functions,

$$r(t) = \sqrt{x(t)^2 + y(t)^2} \quad \text{and} \quad \theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right).$$

In the case of the unit circle parameterized by cosine and sine, these give the constant unit function and the identity map, respectively. For the squircle, however, we have

$$\begin{aligned} r'(t) &= \frac{d}{dt}\sqrt{x(t)^2 + y(t)^2} \\ &= \frac{x(t)x'(t) + y(t)y'(t)}{\sqrt{x(t)^2 + y(t)^2}} \\ &= \frac{-x(t)y(t)^3 + yx(t)^3}{\sqrt{x(t)^2 + y(t)^2}} \\ &= \frac{x(t)y(t)(y(t)^2 - x(t)^2)}{r(t)}. \end{aligned}$$

This illustrates precisely what makes a squircle different from a circle. The zeros of $r'(t)$ correspond to the lines of symmetry of the squircle, with the anticipated sign changes in between. For example, the derivative is positive as we move counterclockwise along the squircle from the point $(1, 0)$, indicating that the squircle is breaking away from the circle. This continues until the intersection with the line $y = x$ at $(\sqrt[4]{1/2}, \sqrt[4]{1/2})$, where the radius begins to decrease until it meets back with the unit circle at (0, 1).

As to the angle,

$$\begin{aligned} \theta'(t) &= \frac{d}{dt}\tan^{-1}\left(\frac{y(t)}{x(t)}\right) \\ &= \frac{1}{1 + \frac{y(t)^2}{x(t)^2}} \cdot \frac{d}{dt}\left(\frac{y(t)}{x(t)}\right) \\ &= \frac{\mathrm{cq}(t)^2}{r(t)^2} \cdot \frac{d}{dt}\mathrm{tq}(t) \quad = \quad \frac{1}{r(t)^2}. \end{aligned}$$

Since the squircle exscribes the circle, we have $r(t) \geq 1$ for all t and so $0 < \theta'(t) \leq 1$, meaning that the resulting parameterization progresses counterclockwise around the unit squircle, but at a slower rate than the angular parameterization familiar in the circle.

We should now observe that there is really nothing special about the squircle parameterization given by (3) except that it was a fairly obvious choice when working backward from the defining equation $x^4 + y^4 = 1$. However, we could have chosen any CIVP of the form

$$\begin{cases} x'(t) = -y(t)^3\,\zeta(x(t), y(t), t) \\ y'(t) = x(t)^3\,\zeta(x(t), y(t), t) \\ x(0) = 1 \\ y(0) = 0 \end{cases} \tag{4}$$

where $\zeta(x(t), y(t), t))$ is any appropriate differentiable function. To see this, suppose u and v satisfy (4) for some function ζ and let $g(t) = u(t)^4 + v(t)^4$. Then $g(0) = 1$ and

$$\begin{aligned} g'(t) &= 4u(t)^3u'(t) + 4v(t)^3v'(t) \\ &= -4u(t)^3v(t)^3\zeta(x(t), y(t), t) + 4u(t)^3v(t)^3\zeta(x(t), y(t), t) = 0 \end{aligned}$$

and so $g(t)$ is identically one, just as we saw earlier for cosquine and squine.

EXERCISE 4. *We saw why the solution curves to a CIVP of the form* (4) *must lie on the squircle, but we want to guarantee that we get the whole squircle. What goes wrong if we let* $\zeta(x, y, t) = x$*? How can we fix this?*

For example, to find the angular parameterization of the squircle we must find the function ζ in (4) so that $\theta'(t) = 1$:

$$\begin{aligned}\theta'(t) &= \frac{1}{1+\frac{y(t)^2}{x(t)^2}} \cdot \frac{d}{dt}\left(\frac{y(t)}{x(t)}\right) \\ &= \frac{x(t)^4\zeta(x(t), y(t), t) + y(t)^4\zeta(x(t), y(t), t)}{x(t)^2 + y(t)^2} = \frac{\zeta(x(t), y(t), t)}{x(t)^2 + y(t)^2},\end{aligned}$$

showing that the CIVP that gives the angular parameterization of the unit squircle is

$$\begin{cases} x'(t) = -y(t)^3\,(x(t)^2 + y(t)^2) \\ y'(t) = x(t)^3\,(x(t)^2 + y(t)^2) \\ x(0) = 1 \\ y(0) = 0. \end{cases} \tag{5}$$

The same procedure can be used to find the arclength parameterization, which, unlike the circle, will be different from the angular parameterization. It is here that we might reflect again on a key difference between the circle and the squircle, which relates to the existence of π. The trigonometric functions are arranged so that their periods reflect simultaneously the arclength *and* the angular parameterizations of the circle. No parameterization of the squircle can do both at the same time. We have values that play some of the roles of π, but none that can do it all.

EXERCISE 5. *Find a CIVP giving the arclength parameterization of the unit squircle.*

EXERCISE 6. *Find the angular parameterization for the unit squircle by expressing the curve in polar coordinates. Compare with* (5).

Inverse functions and identities

If a function is defined as the solution to an initial value problem, it is a simple matter to find a defining IVP for its inverse.

THEOREM 3. *Suppose* $f(t)$ *is differentiable and has an inverse function on an interval containing* t_0. *If* f *satisfies the IVP*

$$\begin{cases} y'(t) = F(y(t), t) \\ y(t_0) = y_0 \end{cases} \tag{6}$$

then the inverse function satisfies

$$\begin{cases} z'(t) = \frac{1}{F(t, z(t))} \\ z(y_0) = t_0. \end{cases} \tag{7}$$

This is just the chain rule. Suppose $f(t)$ satisfies (6) and let $g(t)$ be the inverse function, so that $g(f(t)) = t$ in the specified interval. Then

$$1 = \frac{d}{dt}t = \frac{d}{dt}f(g(t)) = f'(g(t))g'(t) = F(f(g(t)), g(t))g'(t) = F(t, g(t))g'(t)$$

from which it follows that $z = g(t)$ satisfies (7).

We now apply this theorem to develop properties of the inverse squigonometric functions. We have already shown that tanquent satisfies the initial value problem

$$\begin{cases} y'(t) = \sqrt{y(t)^4 + 1} \\ y(0) = 0. \end{cases} \tag{8}$$

Applying Theorem 3, we reciprocate and switch the independent and dependent variables to obtain, by way of the Fundamental Theorem of Calculus, an interesting integration formula involving the inverse:

$$\int_0^t \frac{du}{\sqrt{u^4+1}} = \operatorname{tq}^{-1}(t).$$

Deciding whether this means we have actually "solved" this integral makes for an interesting class discussion.

Other integral formulas such as

$$\int_0^t \frac{du}{(1-u^4)^{3/4}} = \operatorname{sq}^{-1}(t) \tag{9}$$

may also be found with the same trick and some care in restricting domains to properly define the inverses. For this integral, a simple calculation gives $\frac{d}{dt}\operatorname{cq}(\operatorname{sq}^{-1}(t)) = -t^3/(1-t^4)^{3/4}$. This expression can be integrated with a simple substitution, providing the squigonometric Pythagorean identity $\operatorname{cq}(\operatorname{sq}^{-1}(t)) = \sqrt[4]{1-t^4}$.

The identities for squigonemetric functions are similar enough to their trigonometric analogs that we can find relations among them. For example,

$$\frac{d}{dt}\sin^{-1}(\operatorname{sq}(t)^2) = \frac{1}{\sqrt{1-\operatorname{sq}(t)^4}} \cdot 2\operatorname{sq}(t)\operatorname{cq}(t)^3 = 2\operatorname{sq}(t)\operatorname{cq}(t). \tag{10}$$

The integral formula (9) also offers a mechanism for a computer algebra system or those with a penchant for elliptic integrals to express the period of the squine function in terms of the gamma function:

$$\frac{1}{4}\cdot(\text{period of squine}) = \operatorname{sq}^{-1}(1) = \int_0^1 \frac{dt}{(1-t^4)^{3/4}} = \frac{4\Gamma(\frac{5}{4})^2}{\sqrt{\pi}} \approx 1.85407.$$

Indeed, this topic may be taken as invitation to the theory of elliptic integrals. See Exercise 7 and [**5**].

EXERCISE 7. *To work with a more familiar curve, try developing this approach for functions parameterizing the ellipse* $x^2 + 4y^2 = 1$. *Then explore "squircular ellipses."*

EXERCISE 8. *Recover identity* (10) *by directly integrating* $\int \operatorname{sq}(t)\operatorname{cq}(t)\,dt$. *What other identities can you find of this form?*

Generalizations

There are plenty of avenues along which these ideas may be pursued. We point the way on a few of them and invite the reader to develop and extend these ideas in the exercises.

There is nothing really special about the exponent 4 in equation (2), although it is convenient that four is an even integer. In general, the *p-norm* of a 2-vector is the quantity $\|\langle x, y\rangle\|_p = (|x|^p + |y|^p)^{1/p}$. This defines a metric when $p \geq 1$ ([**6**]). The

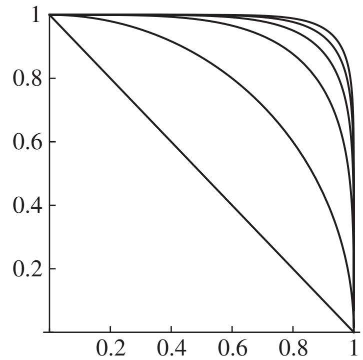


Figure 3 The curves $x^p + y^p = 1$ for $p = 1, 2, 4, 6, 8, 10$

$p = 2$ case, for example, is the familiar euclidean norm $\|\langle x, y\rangle\|^2 = x^2 + y^2$. We have thus been studying circles in the 4-norm and we may expand our definition of a squircle to include the set of points satisfying $|x|^p + |y|^p = 1$ for other values of p. We could generalize our definitions for any squircle by defining $\mathrm{sq}_p(t)$ and $\mathrm{cq}_p(t)$ to be the solutions to the coupled initial value problem

$$\begin{cases} x'(t) = -y(t)^{p-1} \\ y'(t) = x(t)^{p-1} \\ x(0) = 1 \\ y(0) = 0. \end{cases} \tag{11}$$

Again, this parameterization is convenient but basically arbitrary; this and all remaining CIVP's can be generalized like (4).

We can even use the p-norm to define a squircular conformal coordinate system analogous to polar coordinates. The circles in this metric are now squircles, but we can still specify any point in the plane by identifying the squircle on which it lies and how far around to go. As we see in FIGURE 4, the story is a bit more complicated than the polar situation. The curves that meet the squircles at right angles are not straight lines in the euclidean (2-norm) metric and we have already seen that the angular and arclength parameterizations are different.

EXERCISE 9. *Find explicit equations for the curves in* FIGURE 4 *that are orthogonal to the family of concentric squircles.*

EXERCISE 10. *Use the generalized squigonometric functions to find integration formulas and identities for any* $p > 1$.

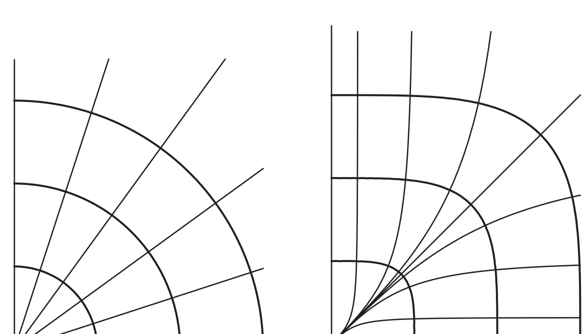


Figure 4 Polar and "squircular" coordinates

EXERCISE 11. *In the terminology of p-norms, we have devoted much of our discussion to comparing trigonometry in the 4-norm with the familiar 2-norm. Another interesting special case is* $p = 1$. *This is the* taxicab norm $\|\langle x, y\rangle\|_T = |x| + |y|$, *so named because it identifies the practical length of a route in a city grid in which a driver will only be able to move on roads running north-south or east-west (i.e., no driving diagonally through buildings!). See* [**4**] *for a delightful exploration of this geometry.*

Use the methods we have discussed to develop trigonometry in the taxicab metric (1-norm) using CIVP's. (Note: So far we have been able to ignore the absolute values in our definition of the p-norm because four is even, but now they become a minor issue. Start by working with $x, y \geq 0$ *and then use symmetry.)*

EXERCISE 12. *Consider the p-norm as p tends to infinity. This leads to the* infinity norm *of a vector, defined as* $\|\langle x, y\rangle\|_\infty = \max(|x|, |y|)$. *This gives a metric whose circles are euclidean squares. Describe* $\mathrm{sq}_\infty(t)$, $\mathrm{cq}_\infty(t)$, *and* $\mathrm{tq}_\infty(t)$.

EXERCISE 13. (Inspired by [**7**]) *Let* $p > 2$ *be an integer and let t be real. Show that* $\mathrm{sq}_p(t)$ *and* $\mathrm{cq}_p(t)$ *cannot both be rational unless one of them is* 0. *(Hint: Use Fermat's Last Theorem.)*

Another generalization is to hyperbolic functions. The familiar $\cosh(t)$ and $\sinh(t)$ may be defined as solutions to the CIVP

$$\begin{cases} x'(t) = y(t) \\ y'(t) = x(t) \\ x(0) = 1 \\ y(0) = 0 \end{cases}$$

and they satisfy the equation of the hyperbola $x^2 - y^2 = 1$. We can see this using the same method we used for the circle. Let $h(t) = \cosh^2(t) - \sinh^2(t)$, so $h'(t) = 2\cosh(t)\sinh(t) - 2\sinh(t)\cosh(t) = 0$ and $h(0) = 1$. It follows that $h(t) = 1$ for all t; that is, $x(t)^2 - y(t)^2 = 1$. This tells us that the curve is a subset of the hyperbola, and the initial condition $h(0) = 1$ shows us that it is the right-hand component.

We may define the squircular counterparts cqh(t) and sqh(t) by the CIVP

$$\begin{cases} x'(t) = y(t)^3 \\ y'(t) = x(t)^3 \\ x(0) = 1 \\ y(0) = 0, \end{cases}$$

or more generally in the p-norm,

$$\begin{cases} x'(t) = y(t)^{p-1} \\ y'(t) = x(t)^{p-1} \\ x(0) = 1 \\ y(0) = 0, \end{cases} \tag{12}$$

Just as squine and cosquine parameterize squared-off circles, the "squinch" and "cosquinch" functions parameterize the squared-off hyperbola $x^4 - y^4 = 1$ (FIGURE 5).

How far can we go? Let's note how far we have come with this idea: we have stumbled onto the field of elliptic integrals, we have found a method to develop trigonometry in alternate metric spaces, and we have a new way to think about parameterizations. We close by recalling that the classical and hyperbolic trigonometric functions are related through complex analysis by $\sin(it) = i\,\sinh(t)$ and $\cos(it) = \cosh(t)$. Although typically verified using Euler's formula $e^{it} = \cos t + i\sin t$, they

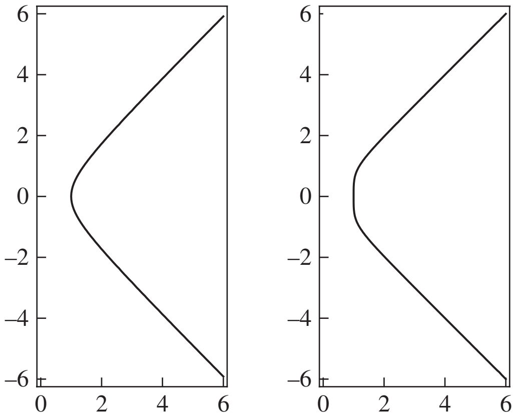


Figure 5 The hyperbola (cosh t, sinh t) and its squircular counterpart (cqh t, sqh t).

may be proved instead using CIVP's. Can we find a similar connection in squigonometry? Our final exercise answers this with a solid "yes."

EXERCISE 14. *Extend the hyperbolic identities* $\sin(it) = i\sinh(t)$ *and* $\cos(it) = \cosh(t)$ *to the p-norm by showing that if* $\omega^p = -1$*, then* $\omega\,\mathrm{sqh}_p(t) = \mathrm{sq}_p(\omega t)$ *and* $\mathrm{cqh}_p(t) = \mathrm{cq}_p(\omega t)$*, where* $\mathrm{cqh}_p(t)$ *and* $\mathrm{sqh}_p(t)$ *are the functions satisfying* (12).

REFERENCES

1. J. Callahan, D. Cox, K. Hoffman, D. O'Shea, H. Pollatsek, and L. Senechal, *Calculus in Context: The Five College Calculus Project*, W.H. Freeman, 1995.
2. B. Cha, Transcendental Functions and Initial Value Problems: A Different Approach to Calculus II, *College Math. J.* **38** (2007) 288–296.
3. W. Boyce and R. DiPrima, *Elementary Differential Equations*, 9th ed., Wiley, 2008.
4. E. F. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*, Dover, 1987.
5. C. C. Maican, *Integral Evaluations Using the Gamma and Beta Functions and Elliptic Integrals in Engineering: A Self-Study Approach*, International Press, 2005.
6. W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, 1986.
7. R. M. Young, *Execursions in Calculus,* Mathematical Association of America, 1992.

Summary Differential equations offers one approach to defining the classical trigonometric functions sine and cosine that parameterize the unit circle. In this article, we adapt this approach to develop analogous functions that parameterize the unit *squircle* defined by $x^4 + y^4 = 1$. As we develop our new theory of "squigonometry" using only elementary calculus, we will catch glimpses of some very interesting and deep ideas in elliptic integrals, non-euclidean geometry, number theory, and complex analysis.

WILLIAM E. WOOD has recently joined the Mathematics Department at the University of Northern Iowa. He enjoys thinking about various problems across mathematics and somehow turning them all into geometry problems. He lives in Cedar Falls with his wife, cats, and board game collection.

The Editor's Song

FRANK A. FARRIS
Santa Clara University
Santa Clara, CA 95053
ffarris@scu.edu

The program of the 2011 Mathfest's opening banquet was "MAA—The Musical!" Produced by Annalisa Crannell and starring the "MAA Players" (active MAA members all), it highlighted activities of the Association and of Mathfest itself. This song represents the journals. It was sung by past editor Frank Farris to the tune of "A Wand'ring Minstrel I," from Gilbert and Sullivan's *the Mikado.*

The Editor:
An editor am I
I hope you read our journals,
we've pithy, mathy kernels
And tales of e and pi.

Our catalog is long
From CMJ to LOCI
The Monthly and singular FOCI,
Mathematics Magazine!
Mathematics Magazine!

Are you inclined to referee
We've work for you!
Oh, bless you!
We need reports in depth, you see!
And on time too.
Oh, bless you, bless you!
We'll use your expertise
in strange geometries.
But pray you keep in mind
Be firm but kind!
Oh bless you, bless you!

And if all electeronic's your selection
You can read the Monthly and the CMJ
And the Magazine and LOCI every section
With only one low fee for you to pay!

Our backlogs all in JSTOR are assembled
You can access them—perhaps for a small fee.
And I shouldn't be surprised if you should stumble
Upon a cache of papers of high pedigree!

Chorus:
We shouldn't be surprised if we should stumble
(and be truly charmed)
Upon a cache of papers of high pedigree!

Math. Mag. **84** (2011) 266–267. doi:10.4169/math.mag.84.4.266.

The Editor: And if an author you would be
Why send your paper round!
We'll log it in and send it off
and have it refereed for free?
Hurrah if the work is sound!

Chorus: Yo ho! Heave ho!
We hope that our work is sound.

The Editor: To see your work come out in print
Will tickle your chair or dean.
But the happiest hour an author sees
Is when she's down
At a coffee shop in town
With her laptop on her knee, yo ho!
And the TEX flowing bold and free.

Chorus: Then let's be authors—off we go!
As the editor drums us round.
With a yeo heave ho,
And a rum below,
We'll hope that our work is sound!

The Editor: An editor am I
I hope you read our journals
with pithy, mathy kernels
And tales of e and pi.
And 3.14159,
e and π.

Frank Farris sings to an adoring cast of MAA Players. Left to Right, these are Jenny Quinn, Talithia Williams, Annalisa Crannell, Jennifer Beineke, Alissa Crans, Norm Richert, Dan Kalman, Matt DeLong, Francis Su, and Art Benjamin's hand. Photo by Laura McHugh.

The Mathematics of Referendum Elections and Separable Preferences

JONATHAN K. HODGE
Grand Valley State University
Allendale, MI 49401
hodgejo@gvsu.edu

Referendum elections, as practiced in the United States, often require voters to register simultaneous yes or no votes on multiple issues or proposals. But what happens when these proposals are related? For instance, consider a situation in which one proposal asks voters to approve of expensive road improvements, while another proposes a gasoline tax that could pay for the improvements. A voter may favor the road improvements, but only if the gas tax is approved, or vice versa. The interdependency of these issues poses a dilemma on voting day, as voters must predict whether one proposal will pass in order to determine how to vote on the other. If voters' predictions are wrong—or if voters simply ignore the underlying interdependence—then undesirable and even paradoxical election outcomes can occur. For instance, it is possible for a particular combination of outcomes to win, even though it is the least-preferred combination of every voter.

Lacy and Niou [**17**] note that although "the resurrection of direct democracy through referendums is one of the clear trends of democratic politics... referendums as currently practiced force people to separate their votes on issues that may be linked in their minds." This so-called *separability problem* is interesting from a practical perspective and also leads to a number of important theoretical questions, many of them mathematical.

In this article, we summarize some recent mathematical contributions to the separability problem and suggest several directions for further research. We begin by formally defining the notion of separability as it pertains to voter preferences in referendum elections. We then show that separable preferences are structurally complex, rare in random electorates, and sensitive to small changes. As it turns out, they are also crucial to the problem of obtaining outcomes in referendum elections that are reasonably reflective of the preferences of the voters.

Separable preferences

Separability has to do with whether a voter's preferences on one or more questions in a referendum election may depend on the known or predicted outcomes of other questions in the election. Preferences that are free from such interdependence are said to be *separable*.

When a voter's preferences are separable, the voter's best strategy (technically, a weakly dominant strategy) is to vote sincerely—that is, for one's most preferred outcome on each proposal [**17**]. For voters whose preferences are not separable, strategic voting (voting for an outcome other than one's first choice) can sometimes yield significantly better outcomes. But strategic voting is far from a perfect solution to the separability problem. For instance, Lacy and Niou [**17**] have shown that when some voters have nonseparable preferences, strategic voting can fail to elect a Condorcet

Math. Mag. **84** (2011) 268–277. doi:10.4169/math.mag.84.4.268.

winner (that is, an outcome that is preferred by a majority to every other possible outcome), even when one exists. Kilgour and Bradley [**16**] note when voters have non-separable preferences, every strategic choice carries with it the risk of *regret*—that is, the realization that a different choice would have yielded a better outcome.

In order to formalize the notion of separability, we must first have a way to model voter preferences in a referendum election. We restrict our attention to elections that consist of a finite number, say n, of yes/no decisions. In this context, an *outcome* is defined to be a vector of 0's and 1's, one for each proposition. In other words, an outcome is a point in binary n-space, $\{0, 1\}^n$. We may also refer to the outcome of a particular set of j questions (with $1 \leq j \leq n$), in which case we mean a point in a corresponding j-dimensional subspace of $\{0, 1\}^n$. We think of a 1 in the kth component of an outcome as representing YES on question k, whereas a 0 represents NO. We then use a total order on $\{0, 1\}^n$ to represent a voter's preferences over all possible outcomes.

A *binary preference matrix* [**3**] provides a convenient way of visualizing a voter's preferences. In lieu of a formal definition, we consider the following example.

Example Let the question set be $Q = \{1, 2\}$. The 4×2 matrix below represents a voter's preferences over the four possible outcomes of an election on the two questions in Q:

$$P_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

The interpretation of this matrix is that an outcome A is preferred to an outcome B if and only if A appears above B in P_1. We then note the following:

- The voter's most preferred outcome is YES on the first question and NO on the second question.
- The voter's least preferred outcome is NO on both questions.
- YES is always preferred to NO on the first question (regardless of the outcome on the second question).
- NO is preferred to YES on the second question, provided that the outcome on the first question is YES. If the outcome of the first question is NO, then this preference is reversed.

Could these preferences occur in practice? In fact, they might be very natural if Proposition 1 calls for building a new elementary school, while Proposition 2 calls for building an addition to an older one.

In this example, we noted that the voter's preference on the first question does not depend on the outcome of the second question. Because of this independence, we say that question 1 is *separable*. More precisely, we say that the set $\{1\}$ is separable with respect to P_1. On the other hand, since the voter's preference on the second question does depend on the outcome of the first question, we say that the set $\{2\}$ is *not* separable with respect to P_1. We adopt the convention that the entire question set, Q, and the empty set, ϕ, are separable with respect to any preference matrix. Thus, we can associate with P_1 the following collection of separable sets, called the *character* of P_1 and denoted by $\text{char}(P_1)$:

$$\text{char}(P_1) = \{\phi, \{1\}, \{1, 2\}\}.$$

The next example further demonstrates the idea of separability and also motivates the precise definition we will adopt shortly.

Example Consider the following preference matrix for an election on 3 questions:

$$P_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Suppose we fix an outcome of 1 on the third question. This choice induces the following order on the possible outcomes of the first and second questions:

$$11 \succ 01 \succ 10 \succ 00$$

If we instead fix 0 on the third question, a different order is induced:

$$11 \succ 10 \succ 01 \succ 00$$

Since the ordering of the outcomes on the first and second questions depends on which choice we fix for the third question, we say that $\{1, 2\}$ is not separable (with respect to P_2).

The two induced orders from this example correspond uniquely to the following submatrices of P_2:

$$P_2^{[\{3\},1]} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_2^{[\{3\},0]} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

The notation we use to describe these submatrices reflects the fact that we are fixing an outcome on a certain subset of Q. In general, if P is a preference matrix for an election on Q, S is a subset of Q, and x is an outcome on S, then we use the notation $P^{[S,x]}$ to denote the *submatrix of P induced by fixing x on S*. The formal definition of separability now follows naturally.

Definition Let P be a preference matrix for an election on Q, and let S be a proper, nonempty subset of Q. We say that S is *separable with respect to P*, or that P is *separable on S*, if for any two outcomes x and y on $Q - S$,

$$P^{[Q-S,x]} = P^{[Q-S,y]}.$$

As noted above, we consider Q and ϕ to be separable with respect to any preference matrix.

Now that we have defined separability, we can also formally define the character of a preference matrix:

Definition Let P be a preference matrix for an election on Q. The *character of P*, denoted $\text{char}(P)$, is the collection of all subsets of Q that are separable with respect to P; that is,

$$\text{char}(P) = \{S \subseteq Q : P \text{ is separable on } S\}.$$

When $\text{char}(P) = \mathcal{P}(Q)$ (the power set of Q), we say that P is *completely separable*. When $\text{char}(P) = \{\phi, Q\}$, we say that P is *completely nonseparable*.

Structural properties

Notions of separability are important in a variety of disciplines, including economics, political science, and operations research. In these contexts, however, the sets of possible outcomes for each question are typically assumed to be intervals of real numbers. Gorman's Theorem [**7**], a foundational result in mathematical economics, states that under this and a few other fairly innocuous assumptions, the property of separability is preserved by most set operations. In particular, if S and T are separable with respect to a given preference matrix, then $S \cup T$, $S \cap T$, $S - T$, $T - S$, and $S \triangle T$ (the symmetric difference of S and T, defined to be $(S \cup T) - (S \cap T)$) are all separable as well.

Somewhat surprisingly, the same cannot be said when working within the context of referendum elections. Of course, the binary outcome spaces inherent to such elections are not intervals, but rather discrete sets. Nevertheless, it is somewhat counterintuitive to think, for instance, that a voter's preferences on two individual questions may not be influenced by the outcome of other questions in the election, yet when the two questions are viewed together, the voter's preferences on the combination may in fact exhibit such dependence.

The preference matrix P_2 from the preceding example illustrates exactly this phenomenon. In that example, we argued that the set $\{1, 2\}$ is not separable with respect to P_2. However, it is easy to verify that both $\{1\}$ and $\{2\}$ are separable with respect to P_2, each with induced order $1 \succ 0$. Note that this same matrix demonstrates the failure of separability to be preserved by relative complements (since $\{1, 2\} = \{1, 2, 3\} - \{3\}$ and both $\{1, 2, 3\}$ and $\{3\}$ are separable) and symmetric differences, since ($\{1, 2\} = \{1, 3\} \triangle \{2, 3\}$ and both $\{1, 3\}$ and $\{2, 3\}$ are separable).

Interestingly enough, separability is preserved by intersections, not only in this example, but in general [**3**]. This fact implies that the character of any binary preference matrix must be closed under intersections. The corresponding inverse problem is even more interesting:

The admissibility problem Let Q be a finite set of questions, and let $\mathcal{C}$ be any collection of subsets of Q that contains both ϕ and Q and is closed under intersections. Does there exist a preference matrix P such that $\text{char}(P) = \mathcal{C}$? (If so, we say that $\mathcal{C}$ is *admissible.*)

The admissibility problem has been investigated for elections with 2, 3, and 4 questions [**13**]. In both the 2 and 3 question cases, the answer is a definitive *yes.*

However, in the case of 4 questions, an interesting anomaly arises. Up to a naturally defined notion of isomorphism, there are 165 distinct collections of subsets of $Q = \{1, 2, 3, 4\}$ that satisfy the conditions specified in the problem. Of these, 164 arise as the character of one or more binary preference matrices. This one, however, does not:

$$\mathcal{C}_* = \{\phi, \{1, 2\}, \{2\}, \{2, 3\}, \{3\}, \{3, 4\}, \{1, 2, 3, 4\}\}$$

One way to prove the indamissibility of $\mathcal{C}_*$ is to show that a preference matrix cannot be separable on all of the sets in $\mathcal{C}_*$, and *only* the sets in $\mathcal{C}_*$. In particular, it can be shown that if the sets $\{1, 2\}$, $\{2\}$, $\{2, 3\}$, $\{3\}$, and $\{3, 4\}$ are all separable with respect to a particular preference matrix, then either $\{1, 2, 3\}$ or $\{2, 3, 4\}$ must also be separable. Since neither of these sets belong to $\mathcal{C}_*$, it is impossible for $\mathcal{C}_*$ to be the character of a preference matrix in a four-question election.

It is of course difficult to generalize from one example, and to date, little progress has been made on the problem of classifying all possible characters for elections with 5 or more questions. This lack of progress is due partially to the large number of characters that are closed under intersections (14480 in the 5 question case) and the even larger number of preference matrices ($2^n!$ for an n question election) that must be considered.

From a practical perspective, it would be easy to dismiss examples like the one above as nothing more than a mathematical curiosity. However, there are practical implications to be found. For instance, the fact that so many characters are admissible suggests that there are many complex ways in which a voter's preferences can be interdependent. On the other hand, the fact that not all closed characters are admissible suggests that there is at least some inherent structure to these interdependencies that goes beyond simple closure under intersections. We have yet to fully understand what this structure is or why it occurs. Furthermore, we do not know which characters are more likely to arise in actual voter preferences; we know only what is theoretically possible or impossible, and this only for small question sets. If we are to have any hope of framing elections in ways that will better account for interdependent preferences, we must first understand the ways in which the underlying interdependencies can arise. As such, more work is needed from both theoretical and empirical perspectives.

Combinatorial questions

Preference separability is a common assumption throughout much of the literature in economics and social choice. Therefore, it is natural to consider how common separable preferences are among all possible preferences.

The task of counting separable preference matrices is a difficult one, and exact results are known only for elections with 7 or fewer questions [**3, 11**]. Other related combinatorial questions, however, have been more definitively answered. For instance, it has been shown [**3**] that every separable preference matrix must be *bitwise symmetric*, meaning that the kth row from the top is the bitwise complement of the kth row from the bottom. (Note that the converse does not hold; we leave the proof of this as an exercise.) It is easy to show that the number of bitwise symmetric preference matrices for an election with n questions is

$$2^n \cdot (2^n - 2) \cdot (2^n - 4) \cdots 4 \cdot 2 = 2^{2^{n-1}} \cdot 2^{n-1}!$$

Thus, the probability of a randomly selected preference matrix being separable is bounded above by

$$\frac{2^n \cdot (2^n - 2) \cdot (2^n - 4) \cdots 4 \cdot 2}{2^n \cdot (2^n - 1) \cdot (2^n - 2) \cdots 2 \cdot 1} = \frac{1}{(2^n - 1) \cdot (2^n - 3) \cdot (2^n - 5) \cdots 3 \cdot 1},$$

which approaches 0 (very quickly!) as $n \to \infty$. Even stronger asymptotic results have been obtained by enumerating *preseparable* and *strongly preseparable* preferences, which can be constructed using lattice paths and counted by the Catalan numbers and central binomial coefficients [**11**]. These methods also reveal interesting and largely unexplored connections to Boolean term orders [**21**] and comparative probability relations [**6**]. Finally, recent results [**13**] establish that the probability of *complete nonseparability* approaches 1 as $n \to \infty$. In particular, the probability of a preference matrix being separable on at least one nontrivial subset of Q (that is, a subset other than ϕ or Q) is bounded above by

$$\frac{1}{2^{2^{n-1}-n-1}}.$$

From a numerical perspective, this result implies that for elections with at least 5 questions, more than 99.9% of all possible preference matrices exhibit no separability whatsoever; that is, the preferences represented are as interdependent as possible.

Stated another way, in random electorates, there are far more ways for a voter's preferences over multiple issues to be highly interdependent than there are ways for them to be separable. Of course, it seems reasonable to assume that not all randomly selected preference matrices are equally likely to correspond to the preferences of voters in actual elections. Perhaps some orders are unrealistic and should be eliminated from consideration. If this is the case, then among all *realistic* preference matrices, however that notion is defined, separable preferences may be more prevalent.

Permutations of separable preference orders

Related to the combinatorial results in the previous section is the question of which permutations, when applied to the rows of a separable preference matrix, will necessarily produce another separable matrix. To formally answer this question, note that for an election with n questions, the symmetric group of degree 2^n, denoted S_{2^n}, acts in a natural way on the rows of any preference matrix. A permutation $\sigma \in S_{2^n}$ is said to *preserve separability* if the image of each separable preference matrix under σ is also separable. An analogous definition applies to permutations that *preserve symmetry*.

It has been shown [**9**] that the sets of symmetry-preserving and separability-preserving permutations are each subgroups of S_{2^n}. Furthermore, for an election with n questions, the group of symmetry-preserving permutations is isomorphic to the group of symmetries of a 2^{n-1} dimensional hypercube, which is exactly the wreath product

$$Z_2 \wr S_{2^{n-1}} = [Z_2]^{2^{n-1}} \rtimes S_{2^{n-1}}.$$

More interesting is the fact that for every $n \geq 4$, the group of separability-preserving permutations is isomorphic to the Klein 4-group, generated by

$$\sigma_1 = (2^{n-1}, 2^{n-1} + 1)$$

and

$$\sigma_2 = (1, 2^n)(2, 2^n - 1) \cdots (2^{n-1}, 2^{n-1} + 1).$$

This means that, for elections with at least four questions, only four permutations preserve the separability of all separable preference matrices. From a practical perspective, this result indicates that the desirable property of separability can be affected by small changes in voter preferences. Separable and nonseparable voter preferences do not need to be drastically different. In fact, a change as simple as the transposition of two adjacent rows can make the difference between separability and nonseparability.

Of course, most voters do not think about changes in their preferences from a group theoretic perspective. But preferences do evolve over time, and the above results suggest that the steps in this evolutionary process can introduce interdependencies that make preferences harder to express in a simultaneous election over multiple issues.

Practical implications of separability

In our first example, we suggested that it is possible for a referendum election to result in an outcome that is viewed as one of the least favorable options by each voter in the electorate. In fact, it is relatively easy to construct an example in which the winning outcome is the least preferred choice of *every* voter. (In the voting theory literature, such an outcome is said to be *universally Pareto-dominated*.) For a three question election, simply assign each of the following preference matrices to an equal number of voters:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

If all voters vote for their first choices (that is, the top rows of their preference matrices) then every question will pass by a two-thirds vote, yielding the outcome (1, 1, 1).

Note that none of these preference matrices is separable. This fact is significant, although it should also be noted that highly undesirable results can occur even when all but one of the voters (in an arbitrarily large electorate) have separable preferences [**10, 17**]. Note also that the outcome illustrated by this example is affected only by the first and last rows of the given matrices. Thus, there are many variations that would produce the same effect.

Some research indicates that election outcomes can be improved significantly if the questions can be framed in such a way that *all* voters have separable preferences [**15, 17, 27**]. This approach, however, seems unrealistic, especially in elections with a large number of proposals. It may be more reasonable to consider whether any incremental gains can be realized if some, but not all, voters have separable preferences. In other words, is some separability better than no separability, and is there a correlation between the relative degree of separability present in an electorate and the desirability of the resulting election outcomes?

Recent computer simulations suggest an affirmative answer to this question [**12**]. In particular, the data obtained from these simulations reveal a linear relationship between the percentage of voters in an election that have separable preferences and the desirability of the resulting outcome, as quantified by a measure called the *aggregate score index* (ASI). The ASI of an outcome is simply the average of the voters' rankings of that outcome, with rankings ranging from 0 (for a bottom-ranked outcome) to $2^n - 1$ (for a top-ranked outcome). (Note the relation to the the well-known Borda count.)

FIGURE 1 illustrates the aforementioned linear relationship for the 3 question case. Note that, in this example, 0 is the lowest possible ASI and 7 is the highest.

The relatively small difference in ASI (3.63 to 3.79) between electorates with no separable preferences and those with all separable preferences might suggest that the effect of separability is minimal. However, if one considers not only the numerical score of each election result, but also how that result compares to other potential outcomes, a much stronger effect emerges, as shown in TABLE 1. (The data are for 10,000 simulated elections, each with 3 questions and 100 voters.)

Note that when all voters have separable preferences, the first or second best outcome (as measured by the ASI) is selected in over 88% of the simulated elections. In contrast, when no voters have separable preferences, the first or second best outcome is selected less than 43% of the time. Equally striking is the fact that the worst or second worst outcome is selected in over 11% of all elections in which voters have nonseparable preferences, as compared to less than 0.5% of elections in which voters' preferences are separable.

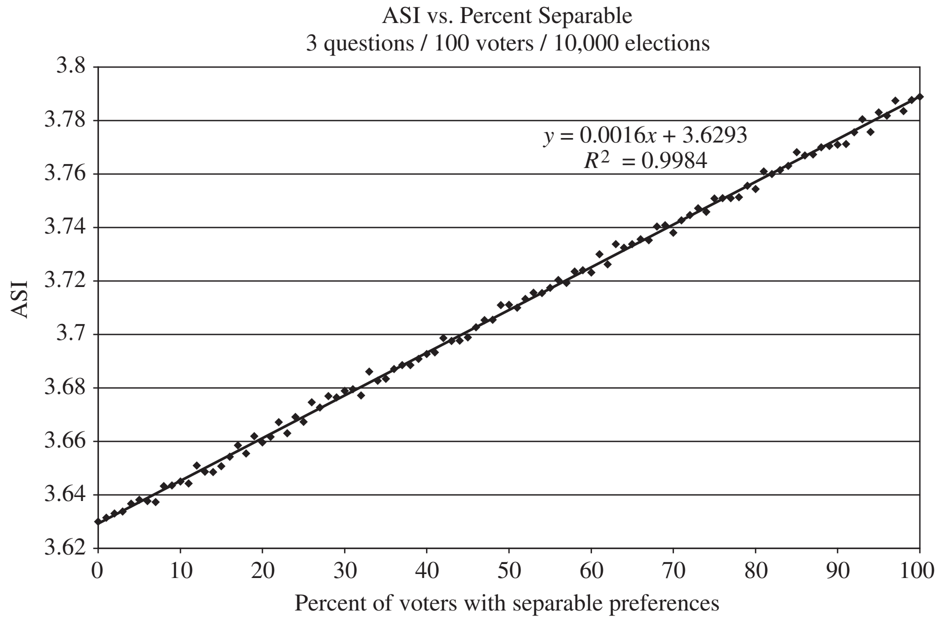


Figure 1 ASI vs. % Separable—Data from 10,000 simulated elections

TABLE 1: The effect of separability on selecting ASI-optimal outcomes

% Sep ↓	Percent of elections resulting in the ith best outcome, where $i = \cdots$							
	1	2	3	4	5	6	7	8
0	25.33	17.53	13.94	12.72	10.26	8.55	7.08	4.59
10	27.78	19.23	14.33	11.84	9.19	8.02	5.49	4.12
20	29.43	19.95	14.99	12.17	8.22	7.20	4.72	3.32
30	31.61	21.88	13.95	11.02	8.57	6.46	4.16	2.35
40	33.64	21.73	14.25	11.49	7.43	5.71	3.69	2.06
50	37.84	23.60	13.25	10.63	5.78	4.05	2.98	1.87
60	40.43	23.83	13.93	9.52	5.31	3.46	2.13	1.39
70	44.52	24.80	13.10	8.59	4.09	2.63	1.35	0.92
80	48.71	26.11	11.67	7.22	2.77	1.72	1.15	0.65
90	55.07	25.35	9.57	5.67	1.86	1.46	0.56	0.46
100	63.72	24.39	5.70	4.28	0.76	0.70	0.26	0.19

Conclusions

In the last decade, insight into the separability problem has been significantly enhanced by the work of mathematicians and mathematical social scientists, who have shown that voter preferences in referendum elections often contain structurally complex interdependencies, and that these interdependencies can substantially impact the how well election outcomes reflect voter preferences. In random electorates, separable preferences, which are the most desirable from an interdependence standpoint, are rare and sensitive to small changes.

Of course, all of these results depend on the particular model used to represent voter preferences, and on the system used to aggregate these preferences and arrive at a collective decision. Other models of voter preference may produce different insights into the structure, prevalence, and importance of separable and nonseparable preferences. Furthermore, there are indications that alternative voting and aggregation methods (such as sequential voting and setwise aggregation [**12**]) may hold some promise for obtaining practical solutions to the separability problem.

Our investigations here have focused primarily on the separability problem as it pertains to referendum elections. For the reader who is interested in exploring this topic in more detail, there are plenty of other interesting results that we have not mentioned here (for instance, [**14, 20**]). Furthermore, separable preferences are important to many other areas and applications, including fair division [**8, 23**], committee selection [**25, 26**], game theory [**1, 2, 5, 22**], artificial intelligence [**18, 19, 28**], and conflict resolution [**24**]. With regard to the latter, Pruitt and Kim [**24**] note that "demands, goals, aspirations, and values often come in bundles—that is, they are psychologically linked" (p. 208), and that, unless these linkages are broken, the most integrative solutions to a conflict situation may remain unexplored. To this date, there seems to be very little published work involving mathematical models of separability within the theory of social conflict.

All of these areas of investigation have the potential to yield rich mathematical results, some of which may have connections or applications to other fields of mathematics. They also provide opportunities for mathematicians to contribute to the ongoing debate about how to best implement the ideals of democracy. There may not be a simple solution to the separability problem, but further mathematical research holds the potential to help minimize its negative effects on the outcomes of referendum elections and other multidimensional decision-making processes.

Acknowledgments Figure 1 appears also as Figure 1 of [**12**], and is reprinted with kind permission from Springer Science+Business Media (Copyright 2006 Springer).

The author wishes to acknowledge Matt Boelkins for his helpful suggestions on earlier drafts of this article.

REFERENCES

1. Andrea Attar, Dipjyoti Majumbar, Gwena el Piaser, and Nicolás Porteiro, Common agency games: Indifference and separable preferences, *Math. Soc. Sci.* **56** (2008), 79–95. http://dx.doi.org/10.1016/j.mathsocsci.2007.11.001
2. Anna Bogomolnaia and Matthew O. Jackson, The stability of hedonic coalition structures, *Games Econ. Behav.* **38** (2002), 201–203. http://dx.doi.org/10.1006/game.2001.0877
3. W. James Bradley, Jonathan K. Hodge, and D. Marc Kilgour, Separable discrete preferences, *Math. Soc. Sci.* **49** (2005), 335–353. http://dx.doi.org/10.1016/j.mathsocsci.2004.08.006
4. Steven J. Brams, D. Marc Kilgour, and William S. Zwicker, Voting on referenda: The separability problem and possible solutions, *Elect. Stud.* **16** (1997), 359–377. http://dx.doi.org/10.1016/S0261-3794(97)00015-2
5. Nadia Burani and William S. Zwicker, Coalition formation games with separable preferences, *Math. Social Sci.* **45** (2003), 27–52. http://dx.doi.org/10.1016/S0165-4896(02)00082-3
6. Terrence Fine and John Gill, The enumeration of comparative probability relations, *Ann. Probab.* **4** (1976), 667–673. http://dx.doi.org/10.1214/aop/1176996036
7. W. M. Gorman, The structure of utility functions, *Rev. of Econ. Stud.* **35** (1968), 367–390. http://dx.doi.org/10.2307/2296766
8. Claus-Jochen Haake, Matthias G. Raith, and Francis Edward Su, Bidding for envy-freeness: A procedural approach to n-player fair-division problems, *Soc. Choice Welf.* **19** (2002), 723–749. http://dx.doi.org/10.1007/s003550100149
9. Jonathan K. Hodge, Permutations of separable preference orders, *Discrete Appl. Math.* **154** (2006), 1478–1499. http://dx.doi.org/10.1016/j.dam.2005.10.015

10. Jonathan K. Hodge and Richard E. Klima, *The Mathematics of Voting and Elections: A Hands on Approach* Mathematical World #22, American Mathematical Society, 2005.
11. Jonathan K. Hodge, Mark Krines, and Jennifer Lahr, Preseparable extensions of multidimensional preferences, *Order* **26** (2009), 125–147. http://dx.doi.org/10.1007/s11083-009-9112-1
12. Jonathan K. Hodge and Peter Schwallier, How does separability affect the desirability of referendum election outcomes? *Theory and Decision* **61** (2006), 251–276. http://dx.doi.org/10.1007/s11238-006-9001-7
13. Jonathan K. Hodge and Micah TerHaar, Classifying interdependence in multidimensional binary preferences, *Math. Soc. Sci.* **55** (2008), 190–204. http://dx.doi.org/10.1016/j.mathsocsci.2007.07.005
14. İpek Özkal Sanver and M. Remzi Sanver, Ensuring pareto optimality by referendum voting, *Soc. Choice Welf.* **27** (2006), 211–219. http://dx.doi.org/10.1007/s00355-006-0101-7
15. Joseph Kandane, On division of the question, *Public Choice*, **13** (1972), 47–54. http://dx.doi.org/10.1007/BF01718851
16. D. Marc Kilgour and W. James Bradley, Nonseparable preferences and simultaneous elections. Paper presented at American Political Science Association, Boston, MA, Sept. 1998.
17. Dean Lacy and Emerson M.S. Niou, A problem with referendums, *J. Theor. Polit.* **12** (2000), 5–31. http://dx.doi.org/10.1177/0951692800012001001
18. Jérôme Lang, Vote and aggregation in combinatorial domains with structured preferences. In *Proceedings of IJCAI-2007*, 2007.
19. Jérôme Lang and Jérôme Mengin, The complexity of learning separable *ceteris paribus* preferences. In *Proceedings of IJCAI-2009*, 2009.
20. Jérôme Lang and Lirong Xia, Sequential composition of voting rules in multi-issue domains, *Math. Social Sci.* **57** (2009), 304–342. http://dx.doi.org/10.1016/j.mathsocsci.2008.12.010
21. Diane Maclagan, Boolean term orders and the root system B_n, *Order* **15** (1999), 279–295. http://dx.doi.org/10.1023/A:1006207716298
22. Igal Milchtaich, Weighted congestion games with separable preferences, *Games Econ. Behav.* **67** (2009), 750–757. http://dx.doi.org/10.1016/j.geb.2009.03.009
23. Hervé Moulin, *Axioms of Cooperative Decision Making*, Econometric Society Monographs No. 15., Cambridge University Press, New York, 1988.
24. Dean Pruitt and Sung Hee Kim, *Social Conflict: Escalation, Stalemate, and Settlement*, 3rd ed., McGraw-Hill, New York, 2004.
25. Thomas C. Ratliff, Some startling inconsistencies when electing committees, *Soc. Choice Welf.* **21** (2003), 433–454. http://dx.doi.org/10.1007/s00355-003-0209-y
26. Thomas C. Ratliff, Selecting committees, *Public Choice* **126** (2006), 343–355. http://dx.doi.org/10.1007/s11127-006-1747-5
27. Thomas Schwartz, Collective choice, separation of issues, and vote trading, *American Political Science Review*, **71** (1977), 999–1010. http://dx.doi.org/10.2307/1960103
28. Lirong Xia, Vincent Conitzer, and Jérôme Lang, Voting on multiattribute domains with cyclic preference dependencies. In *Proceedings of the Twenty-Third Conference on Artificial Intelligence*, 2008.

Summary Voters in referendum elections are often required to cast simultaneous ballots on several possibly related questions or proposals. The separability problem occurs when a voter's preferences on one question or set of questions depend on the known or predicted outcomes of other questions. Nonseparable preferences can lead to seemingly paradoxical election outcomes, such as a winning outcome that is the last choice of every voter. In this article, we survey recent mathematical results related to the separability problem in referendum elections. We explore the structure of interdependent preferences, consider related combinatorial and algebraic results, and examine the practical impact of separability on the outcomes of referendum elections.

JONATHAN K. HODGE is an Associate Professor of Mathematics at Grand Valley State University in Allendale, MI. He is a graduate of Calvin College (B.S., 1998) and Western Michigan University (Ph.D., 2002) and is currently working toward an M.A. degree in negotiation, conflict resolution, and peacebuilding from California State University-Dominguez Hills. Prof. Hodge is the co-PI of Grand Valley State University's summer mathematics REU, a position that allows him to involve undergraduates in much of his research.

NOTES

How Cinderella Won the Bucket Game (and Lived Happily Ever After)

ANTONIUS J.C. HURKENS
5443 NS Haps, Netherlands
hurkens@science.ru.nl

COR A.J. HURKENS
University of Eindhoven
5600 MB Eindhoven, Netherlands
wscor@win.tue.nl

GERHARD J. WOEGINGER
University of Eindhoven
5600 MB Eindhoven, Netherlands
gwoegi@win.tue.nl

Five identical empty buckets of capacity b stand in the corners of a regular pentagon. Cinderella and her wicked Stepmother play a game that goes through a sequence of rounds: at the beginning of every round, the Stepmother takes one gallon of water from the nearby river, and distributes it arbitrarily among the five buckets. Then Cinderella chooses a pair of neighboring buckets, empties them into the river, and puts them back into the pentagon. Then the next round begins.

The Stepmother's goal is to make one of these buckets overflow. Cinderella's goal is to prevent this. For which bucket sizes b can the Stepmother eventually force a bucket overflow? For which bucket sizes can Cinderella manage to keep the game running forever?

Here are two quick and straightforward observations. First, for buckets of size $b < 1$ the wicked Stepmother can win the game right away by pouring her entire gallon into a single bucket. Second, Cinderella can easily win the game for buckets of size $b \geq 5$ by cycling through the buckets so that each bucket is emptied every five turns (and this even works if she only empties one bucket per round).

How do we go beyond these observations? What's going on for bucket sizes b in the interval $1 \leq b < 5$? To answer these questions, we will have to design appropriate strategies for Cinderella, and we will also have to understand the corresponding counter-strategies for the Stepmother. We need to characterize critical situations that may show up during the game. A crucial step in our analysis is understanding properties and conditions that are invariant and that hence remain valid throughout the game. In the end, it will turn out that the most intuitive strategies are not necessarily the best ones.

Some notation: throughout we denote the five buckets by B_1, B_2, B_3, B_4, B_5, where bucket B_i is adjacent to B_{i-1} and B_{i+1}. All indices are taken modulo 5, so that $B_i = B_{i+5}$; the same convention applies to all other variables dealing with the buckets.

Math. Mag. **84** (2011) 278–283. doi:10.4169/math.mag.84.4.278.

Strategies for Cinderella

We observed above that Cinderella wins the game for $b \geq 5$ by cycling through the buckets, even if she only empties one bucket per round. But why should Cinderella waste half of every move? A more effective strategy is ROUND ROBIN: in the first round empty buckets B_1 and B_2, in the second round B_3 and B_4, in the next round B_5 and B_1, next B_2 and B_3, then B_4 and B_5, and so on. ROUND ROBIN is computationally efficient, as it is independent of the Stepmother's choices and the contents of the buckets. Since ROUND ROBIN empties each bucket every two or three turns, Cinderella wins for $b \geq 3$. And for bucket sizes $b < 3$ the Stepmother wins by putting her first three gallons into the last bucket B_5. We summarize:

For buckets of size $b \geq 3$, ROUND ROBIN keeps the game running forever.
For buckets of size $b < 3$, ROUND ROBIN can be forced into an overflow.

If you think about the bucket problem for five more minutes, then you will probably detect the so-called GREEDY strategy, which shortsightedly tries to make the most obvious progress and which is the most natural approach for engineers: *"Always remove as much water as possible from the system."* More precisely, GREEDY will always empty the pair of neighboring buckets with the largest total contents. By definition, this GREEDY strategy is an excellent short-term strategy. We will show that also its long-term behavior is not too bad:

For buckets of size $b \geq 17/8$, GREEDY keeps the game running forever.
For buckets of size $b < 17/8$, GREEDY can be forced into an overflow.

We postpone the (somewhat technical) analysis of GREEDY to the second half of the paper, because there is another strategy SIMPLE for Cinderella that is extremely primitive, very easy to describe, and better than GREEDY. SIMPLE is built around the following invariant (I.0) that must hold at the beginning of every round.

(I.0) Each pair of non-adjacent buckets has total contents at most 1.

Here is the analysis of SIMPLE. Invariant (I.0) holds at the beginning of the first round, when all the buckets are empty. Assume that SIMPLE manages to maintain (I.0) until the beginning of round k. Denote by x_i ($i = 1, 2, 3, 4, 5$) the contents of bucket B_i at the beginning of this round, and denote by y_i the corresponding contents after the Stepmother has distributed her gallon of water.

At the beginning of the round some pair of adjacent buckets, say B_2 and B_3, are empty. This means $x_2 = x_3 = 0$, and since the Stepmother adds only one gallon per round

$$y_2, y_3 \leq 1. \tag{1}$$

The invariant yields $x_1 + x_4 \leq 1$, which together with $x_2 = x_3 = 0$ implies

$$y_1 + y_2 + y_3 + y_4 \leq x_1 + x_2 + x_3 + x_4 + 1 \leq 2. \tag{2}$$

Inequality (2) implies that $y_1 + y_3 \leq 1$ or $y_2 + y_4 \leq 1$. If $y_1 + y_3 \leq 1$, then $y_2 \leq 1$ by (1), and SIMPLE maintains the invariant by emptying the two neighboring buckets B_4 and B_5. If $y_2 + y_4 \leq 1$, then $y_3 \leq 1$ by (1), and SIMPLE maintains the invariant by emptying the two neighboring buckets B_1 and B_5.

To summarize, at the beginning of every round invariant (I.0) is satisfied, and hence every bucket contains at most 1. Since the Stepmother can add at most one further gallon to this, we arrive at our main result for Cinderella:

For buckets of size $b \geq 2$, SIMPLE keeps the game running forever.

Strategies for the Stepmother

Let us switch to the Stepmother's point of view, and consider the BALANCE-5 strategy: *"Distribute the gallon so that all five buckets are filled to the same level."* Hence in round k the Stepmother fills all buckets to the same level a_k, then Cinderella empties two buckets, then in round $k+1$ the Stepmother fills all buckets to level $a_{k+1} = \frac{1}{5}(3a_k + 1)$, and so on. Routine arguments show that for the starting value $a_0 = 0$ the resulting sequence a_k is strictly increasing and converges to the limit $1/2$.

Now assume that the bucket size is $b = 3/2 - \varepsilon$, where $\varepsilon > 0$ is some small real number. The Stepmother applies BALANCE-5 and waits until all buckets are filled to a level $a_k > 1/2 - \varepsilon$. Then Cinderella empties two buckets, and the Stepmother pours her entire next gallon into a non-empty bucket, which brims over.

For buckets of size $b < 3/2$, BALANCE-5 eventually forces an overflow.

A slight modification of BALANCE-5 yields the stronger BALANCE-2 strategy which concentrates entirely on two non-adjacent buckets: *"Distribute the gallon so that buckets B_1 and B_3 both are filled to the same level."* In round k the Stepmother fills both buckets B_1 and B_3 to level a'_k, then Cinderella empties at most one of them, then the Stepmother fills both buckets to a level of at least $a'_{k+1} = \frac{1}{2}(a'_k + 1)$, and so on. For the starting value $a'_0 = 0$, this sequence a'_k is increasing and converges to the limit 1.

Assume that the bucket size is $b = 2 - \varepsilon$ with $\varepsilon > 0$. Eventually BALANCE-2 will reach a level $a'_k > 1 - \varepsilon$, and at the end of this round Cinderella must leave at least one of B_1 and B_3 containing at least a'_k, and the Stepmother then wins the game by adding her next gallon to that bucket.

For buckets of size $b < 2$, BALANCE-2 eventually forces an overflow.

This essentially completes our answer to the Cinderella puzzle: for bucket sizes $b < 2$ the Stepmother can force a bucket overflow by applying BALANCE-2, whereas for bucket sizes $b \geq 2$ Cinderella manages to keep the game running forever by applying the SIMPLE strategy. It remains to give the analysis of GREEDY that we postponed.

Cinderella's analysis of the GREEDY strategy

Now let us analyze Cinderella's GREEDY strategy (which always empties the pair of neighboring buckets with the largest total contents). We will show that GREEDY always ensures the following three invariants at the beginning of a round:

(I.1) The total contents of all five buckets is at most $3/2$.

(I.2) Each single bucket contains at most $9/8$.

(I.3) Each pair of non-adjacent buckets has total contents at most $5/4$.

These three invariants hold at the beginning of the first round, when all buckets are empty. Denote by x_i ($i = 1, 2, 3, 4, 5$) the contents of bucket B_i at the beginning of round k, denote by y_i the corresponding contents after the Stepmother has distributed her gallon of water, and denote by z_i the corresponding contents after GREEDY has made its move. We assume inductively that the values x_i satisfy the three invariants, and we will prove that then also the resulting values z_i satisfy the invariants. Without loss of generality GREEDY empties buckets B_4 and B_5, so that $z_1 = y_1$, $z_2 = y_2$, $z_3 = y_3$, and $z_4 = z_5 = 0$.

We first formulate five useful inequalities (3)–(7), and then we show that the invariants are indeed maintained. Since the Stepmother adds only one gallon to the system, we deduce from (I.1) that

$$y_1 + y_2 + y_3 + y_4 + y_5 \leq 5/2. \tag{3}$$

In a similar way the inequalities $x_i + x_{i+2} \leq 5/4$ in (I.3) yield

$$y_i + y_{i+2} \leq 9/4 \quad \text{for} \quad i = 1, 2, 3, 4, 5. \tag{4}$$

Since GREEDY decides to empty buckets B_4 and B_5, we derive four greedy inequalities $y_1 + y_2 \leq y_4 + y_5$; $y_2 + y_3 \leq y_4 + y_5$; $y_3 + y_4 \leq y_4 + y_5$; and $y_5 + y_1 \leq y_4 + y_5$. The first one of these greedy inequalities implies

$$y_2 \leq y_4 + y_5, \tag{5}$$

and the last two are equivalent to

$$y_3 \leq y_5 \quad \text{and} \quad y_1 \leq y_4. \tag{6}$$

The sum of all four greedy inequalities yields

$$2(y_1 + y_2 + y_3) \leq 3(y_4 + y_5). \tag{7}$$

Now let us turn to the three invariants. From (7) and (3) we conclude that GREEDY maintains invariant (I.1), since the bucket contents z_i (after GREEDY has moved) satisfy

$$z_1 + z_2 + z_3 + z_4 + z_5 = y_1 + y_2 + y_3 \ \leq \ 3/5(y_1 + y_2 + y_3 + y_4 + y_5) \ \leq \ 3/2. \tag{8}$$

What about the second invariant? Both buckets B_4 and B_5 are empty, and hence trivially satisfy $z_4, z_5 \leq 9/8$. Furthermore (6) and (4) imply

$$z_3 = y_3 \leq 1/2(y_3 + y_5) \leq 9/8. \tag{9}$$

A symmetric argument yields $z_1 \leq 9/8$. Hence it remains to deal with bucket B_2. If GREEDY emptied bucket B_2 in the preceding round $k - 1$, then $x_2 = 0$ and $z_2 = y_2 \leq 1$, and we are done. If GREEDY did not empty bucket B_2 in the preceding round, then it must have emptied one or both of B_4 and B_5; by symmetry we assume that this was bucket B_5 and hence $x_5 = 0$. Invariant (I.3) gives us $x_2 + x_4 \leq 5/4$, and combining this with (5) yields

$$y_2 \leq 1/2(y_2 + y_4 + y_5) \leq 1/2(x_2 + x_4 + x_5 + 1) \leq 9/8. \tag{10}$$

This concludes the proof of invariant (I.2), and we move on to invariant (I.3). Since $z_1, z_2, z_3 \leq 9/8$ by (I.2) and since $z_4 = z_5 = 0$, we only need to look into the bucket pair B_1 and B_3. From (6) and (3) we derive

$$z_1 + z_3 = y_1 + y_3 \leq 1/2(y_1 + y_3 + y_4 + y_5) \leq 5/4. \tag{11}$$

To summarize, GREEDY maintains all three invariants (I.1)–(I.3). Since by invariant (I.2) GREEDY never leaves a bucket with contents above 9/8, Cinderella wins the game for buckets of size $b = 17/8$.

The Stepmother's analysis of the GREEDY strategy

Let us assume that Cinderella sticks to this GREEDY strategy. How should the wicked Stepmother behave to make GREEDY look really bad?

Assume that $b < 17/8$, and pick a real number $\varepsilon > 0$ so that $b < 17/8 - 2\varepsilon$. Consider the following two-phase counter-strategy for the Stepmother. During the first phase the Stepmother executes the BALANCE-5 strategy until all buckets are filled to level $a_k > 1/2 - \varepsilon$. The second phase consists of four additional rounds that are listed in TABLE 1. The last line of the table demonstrates that the Stepmother can raise the contents of one of the buckets to $17/8 - 2\varepsilon$. Hence for $b < 17/8$, the Stepmother can win the game against GREEDY. This completes the analysis of GREEDY.

It is instructive to go through Table 1, and to find the point where SIMPLE and GREEDY start making different choices.

TABLE 1: Stepmother versus GREEDY: the four rounds in the second phase.

	B_1	B_2	B_3	B_4	B_5
Stepmother leaves	$1/2-\varepsilon$	$1/2-\varepsilon$	$1/2-\varepsilon$	$1/2-\varepsilon$	$1/2-\varepsilon$
GREEDY moves to	$1/2-\varepsilon$	$1/2-\varepsilon$	$1/2-\varepsilon$	0	0
Stepmother adds	0	$1/4+\varepsilon$	ε	$3/4-2\varepsilon$	0
and thus leaves	$1/2-\varepsilon$	$3/4$	$1/2$	$3/4-2\varepsilon$	0
GREEDY moves to	$1/2-\varepsilon$	0	0	$3/4-2\varepsilon$	0
Stepmother adds	$5/8$	0	0	$3/8$	0
and thus leaves	$9/8-\varepsilon$	0	0	$9/8-2\varepsilon$	0
GREEDY moves to	0	0	0	$9/8-2\varepsilon$	0
Stepmother adds	0	0	0	1	0
and thus leaves	0	0	0	$17/8-2\varepsilon$	0

Final remarks

Cinderella's bucket problem [**2**] was proposed (but not selected) for the 50th International Mathematical Olympiad that took place in Germany in summer 2009. The problem has a serious background [**1**] in data management environments and in multiplexing of tasks in communication networks. Consider for instance a network switch that connects several input ports to a single output channel. Each input port is equipped with a buffer of limited capacity, and data packets arrive online and can be stored in the buffers. In each time slot the switch can transmit packets from some of the buffers to the output channel, and all decisions must be based on the current state without knowledge of future events. We stress that in practice, greedy strategies are very important because they are fast and use little extra memory. In our game, Cinderella's buckets model the buffers, the water models the data that arrives continuously over time, and Cinderella models a scheduling procedure that tries to keep the system stable by preventing buffer overflows.

Finally let us make the connection to the standard language of game theory [**3**]. If we change the game so that the Stepmother scores points according to the maximum amount she can ever get into a bucket, then the game has a value and optimal strategies. (Actually the Stepmother only has a sequence of strategies, whose results converge; this often happens with non-compact strategy spaces.) We have established that the value of this game is 2, and that in the case where Cinderella follows the sub-optimal GREEDY strategy, the Stepmother can achieve a score of 17/8.

REFERENCES

1. M. Chrobak, J. Csirik, Cs. Imreh, J. Noga, J. Sgall, and G.J. Woeginger, The buffer minimization problem for multiprocessor scheduling with conflicts, *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP'2001)*, LNCS 2076, Springer Verlag, 2001, 862–874.
2. G.J. Woeginger, Combinatorics problem C5, *Problem Shortlist of the 50th International Mathematical Olympiad*, Bremen, Germany, 2009, 33–35.
3. M.J. Osborne and A. Rubinstein, *A course in game theory*, 1994, MIT Press.

Summary The paper investigates a combinatorial two-player game, in which one player (Cinderella) wants to prevent overflows in a system with five water-buckets whereas the other player (the wicked Stepmother) wants to cause such overflows. Several sub-optimal and optimal strategies for both players are analyzed in detail.

Edge Tessellations and Stamp Folding Puzzles

MATTHEW KIRBY
Hershey High School
Hershey, PA 17033
maskirby@gmail.com

RONALD UMBLE
Millersville University of Pennsylvania
Millersville, PA 17551
ron.umble@millersville.edu

Which polygons generate a tiling of the plane when reflected in their edges? The complete answer was discovered by Millersville University students Andrew Hall, Joshua York, and the first author in the spring of 2009, and we present it here as a theorem.

THEOREM 1. *A polygon generating a tiling of the plane when reflected in its edges is one of the following eight types: a rectangle; an equilateral,* 60*-right, isosceles right, or* 120*-isosceles triangle; a* 120*-rhombus; a* 60-90-120 *kite; or a regular hexagon.*

Math. Mag. **84** (2011) 283–289. doi:10.4169/math.mag.84.4.283.

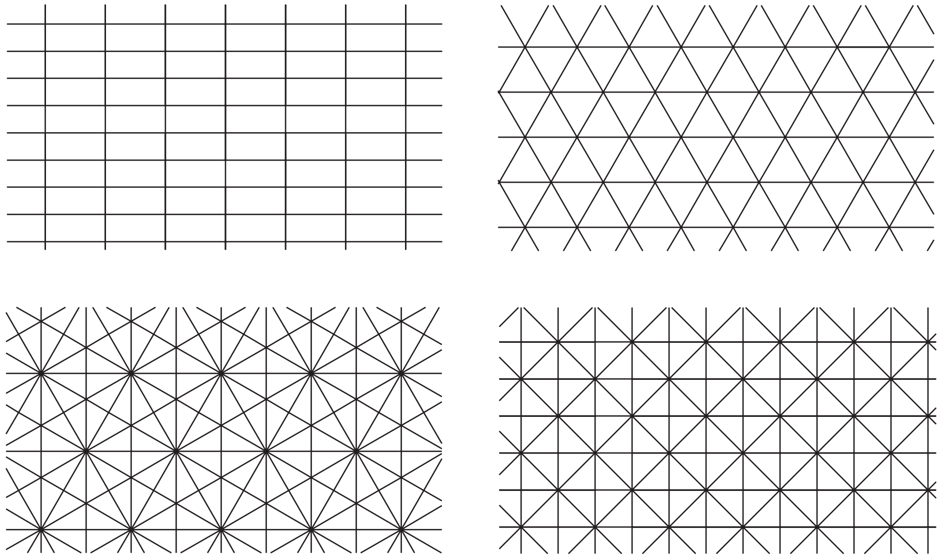


Figure 1 Edge tessellations generated by non-obtuse polygons

A *tessellation* (or *tiling*) of the plane is a collection of plane figures that fills the plane with no overlaps and no gaps. An *edge tessellation* is generated by reflecting a polygon in its edges. The eight edge tessellations, which are pictured in FIGURES 1 and 2, are the most symmetric examples of *Laves tilings*. A discussion of Laves tilings and a complete list of them are given by Grünbaum and Shephard [**4**].

Edge tessellations provide the setting for *stamp folding puzzles*, which are paper folding problems constrained to the perforations on a sheet of postage stamps. The sheet must embed in an edge tessellation, may have any shape, and may be bounded or unbounded, with bounding edges along perforations as in FIGURE 3. The object of a stamp folding puzzle is to create some specified configuration by folding the sheet along its perforations without creasing the stamps. Tucks, which slip one subpacket of

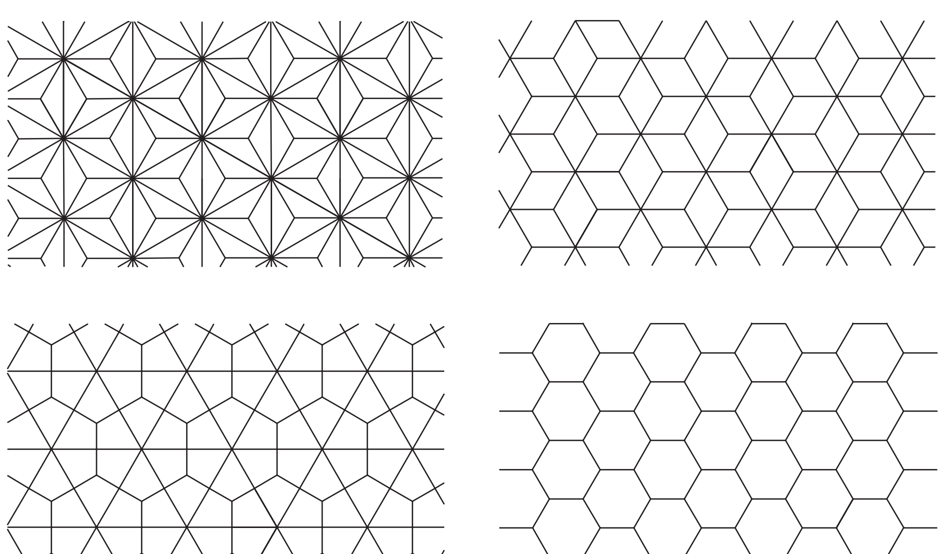


Figure 2 Edge tessellations generated by obtuse polygons

folded stamps between the leaves of another, are allowed. Indeed, a tuck is required to solve the following delightful problem posed by G. Frederickson in his book *Piano-Hinged Dissections: Time to Fold!* [**3**, p. 144]:

> Consider the block of sixteen isosceles right triangular stamps pictured in FIGURE 3. Fold the block into a packet sixteen-deep so that the stamps are arranged in the order 4, 1, 16, 6, 5, 15, 14, 8, 7, 13, 11, 12, 2, 3, 9, 10.

A sheet of postage stamps is suitable for stamp folding puzzles if the stamps are configured in such a way that the sheet folds neatly into a packet of single stamps. Such a packet unfolds into an edge tessellation in which the perforations form lines of symmetry. Frederickson poses the following conjecture [**3**, p. 143]:

> Although triangular stamps have come in a variety of different triangular shapes, only three shapes seem suitable for [stamp] folding puzzles: equilateral, isosceles right, and 60°-right triangles.

Theorem 1 confirms Fredickson's conjecture; indeed, the four suitable edge tessellations are pictured in FIGURE 1. We invite the reader to reproduce and fold each of them into a packet of single stamps. Our folding algorithms appear at the end of this article.

The question posed at the outset of this article and answered by Theorem 1, was motivated by the "unfolding technique" applied by A. Baxter and the second author to find, classify and count classes of periodic orbits of a billiard ball in motion on an equilateral triangular billiard table (see [**2**] for details). Periodic orbits on polygonal billiard tables of the eight polygonal types in Theorem 1 unfold as straight line segments in an edge tessellation. During an REU in 2001, Andrew Baster and students Ethan McCarthy and Jonathan Eskreis-Winkler applied this unfolding technique to find, classify, and count classes of periodic orbits on square, rectangular, and isosceles right triangular billiard tables (see [**1**]). Presumably, this technique also applies on billiard tables of the five remaining polygonal types.

Edge tessellations are *wallpaper patterns*, which are tessellations of the plane with translational symmetries of minimal length in two independent directions (the group of symmetries is discrete). A point C in a wallpaper pattern is an *n-center* if the group of rotational symmetries centered at C is generated by a rotation of minimal positive rotation angle $\phi_n = 360°/n$.

The students' original proof of Theorem 1 applies the powerful *Crystallographic Restriction Theorem*, which tightly constrains the order of a group of rotational symmetries: *If C is an n-center of a wallpaper pattern, then* $n \in \{2, 3, 4, 6\}$ (for a proof,

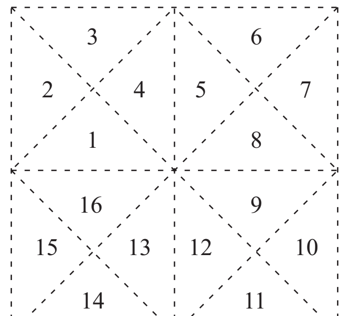


Figure 3 A block of sixteen isosceles right triangular stamps

see [**5**] for example). The proof of Theorem 1 presented here is independent of Crystallographic Restriction and more geometrically revealing.

Proof. We begin the proof of Theorem 1 by constructing a set S containing the measures of the interior angles of a generating polygon G. Let V be a vertex of G, and let θ be the measure of the interior angle at V; then $\theta < 180°$. Let G' be the image of G when reflected in an edge of G containing V. Then the interior angle of G' at V has measure θ, and inductively, the interior angle at V of every copy of G with vertex V has measure θ (see FIGURE 4). Since successively reflecting in the edges of G that meet at V is a rotational symmetry of angle 2θ, the vertex V is an n-center for some n. If G' is the rotational image of G, then $\phi_n = \theta$; otherwise $\phi_n = 2\theta$. In either case, $n\theta = 360°$ for some $n \in \mathbb{N}$, and it follows that every interior angle of G lies in the set

$$\begin{aligned} S &= \{x \leq 120° \mid nx = 360°, n \in \mathbb{N}\} \\ &= \left\{120°, 90°, 72°, 60°, 51\tfrac{3}{7}°, 45°, 40°, 36°, \ldots, 18°, \ldots\right\}. \end{aligned}$$

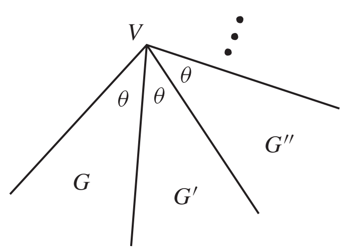


Figure 4 Congruent interior angles at vertex V shared by G and images G' and G''

Now suppose that $\theta = 120°$; then three copies of G share the vertex V. Let e and e' be the edges of G that meet at V, and labeled so that the angle from e to e' measures $120°$ (see FIGURE 5). Let e' and e'' be their respective images under a $120°$ rotation. Then e'' lies on the bisector of $\angle V$ and is the reflection of e' in e. By a similar argument, if an odd number of copies of G share vertex V, the bisector of $\angle V$ is a line of symmetry.

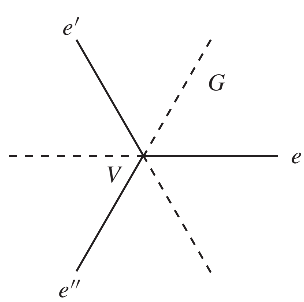


Figure 5 A line of symmetry bisects a $120°$ interior angle

Let g be the number of edges of G. Then the interior angle sum $180°(g-2) \leq 120°g$ implies $g \leq 6$. If G is a hexagon, it is equiangular since its interior angles measure at most $120°$ and its interior angle sum is $720° = 6(120°)$. But G is symmetric with respect to each of its interior angle bisectors by the remark above. Therefore G is a *regular hexagon*.

We claim that G is not a pentagon. On the contrary, suppose G is a pentagon. Then some interior angle measures $120°$ since the interior angle sum of $540° > 5\,(90°)$. Choose an interior angle of $120°$ and label the vertex at this interior angle V. Then G is symmetric with respect to the angle bisector at V and the other interior angles of G pair off congruently—two with measure x, two with measure y. Note that the interior angles in one of these pairs are adjacent (see FIGURE 6). If $x = y$, then $x = 105° \notin S$; hence $x \neq y$. If $x < y$, then $y > 105°$; hence $y = 120°$ since $y \in S$. But if $y = 120°$, lines of symmetry bisect three interior angles of G, in which case $x = y$ by the adjacency noted above, and G is equiangular with an interior angle sum of $600°$, which is a contradiction.

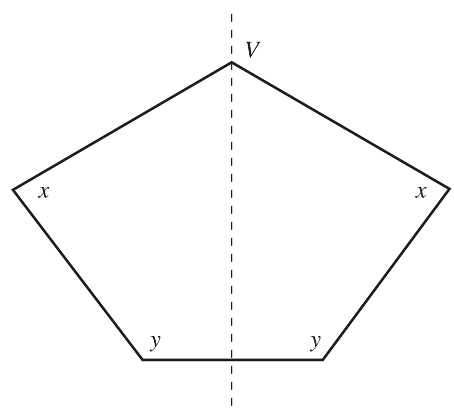


Figure 6 A pentagon G with an interior angle of 120° at V

Now if V is a vertex of G, let $m\angle V$ denote the measure of the interior angle at V. Suppose G is a quadrilateral. If G has an interior angle of $120°$, label the vertices A, B, C, D in succession with $m\angle A = 120°$. Then the bisector s of $\angle A$ is a line of symmetry, C is on s, and $\angle B \cong \angle D$ (see FIGURE 7). Let $2x = m\angle C$ and $y = m\angle B$, and note that $m\angle BAC = 60°$. Then $x \leq 60° \leq 120° - x = y$. Hence the only solutions of $x + y = 120°$ with $x \leq y$ and $(x, y) \in S \times S$ are $\{(30°, 90°), (60°, 60°)\}$. Therefore G is either a 120-*rhombus* or a 60-90-120 *kite*. On the other hand, if the interior angles of G measure at most $90°$, then G is equiangular since its interior angle sum is $360°$, and G is a *rectangle*.

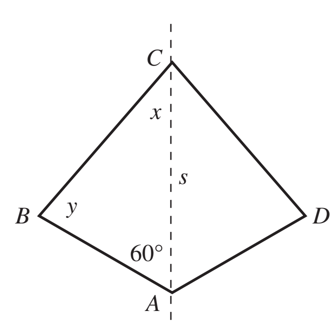


Figure 7 A quadrilateral G with an obtuse interior angle

Finally, suppose G is a triangle. If G has an interior angle of $120°$, then G is a *120-isosceles triangle* by symmetry. Otherwise, let $G = \Delta ABC$; let $x = m\angle A$, $y = m\angle B$, and $z = m\angle C$.

If G is a right triangle with $z = 90°$ and $x \leq y$, then $x, y \in S$ implies $18° \leq x \leq y \leq 72°$. Hence the only solutions of $x + y = 90°$ with $x \leq y$ and $(x, y) \in S \times S$ are $\{(18°, 72°), (30°, 60°), (45°, 45°)\}$. Furthermore, if $y = 72°$, five copies of G share vertex B and the bisector of $\angle B$ is a line of symmetry, in which case $x = y = 90°$, which is a contradiction. Therefore G is either a *60-right* or an *isosceles right triangle*.

If G is an acute triangle with $x \leq y \leq z \leq 72°$, then $x = 180° - (y + z) \geq 180° - 2(72°) = 36°$; on the other hand, $x = 180° - (y + z) \leq 180° - 2(60°) = 60°$. But $36° \leq x \leq 60°$ implies $120° \leq y + z \leq 144°$. Thus if $y \leq z$ and $(y, z) \in S \times S$, then $(y, z) \in \{(60°, 60°), (60°, 72°), (72°, 72°)\}$ so that the only solutions of $x + y + z = 180°$ with $x \leq y \leq z$ and $(x, y, z) \in S \times S \times S$ are $\{(36°, 72°, 72°), (60°, 60°, 60°)\}$. But interior angles of $72°$ are bisected by lines of symmetry, so the solution $(36°, 72°, 72°)$ is extraneous and G is an *equilateral triangle.*

This completes the proof of Theorem 1. ■

We remark that edge tessellations represent 3 of the 17 symmetry types of wallpaper patterns. Using the labeling defined in [**5**], general (non-square) rectangles generate patterns of type *pmm*; isosceles right triangles and squares generate patterns of type *p*4*m*; and the other six polygons in Theorem 1 generate patterns of type *p*6*m*.

Here are some explicit algorithms for folding the sheets of stamps in FIGURE 1 into packets of single stamps. Let T be generated by one of the four non-obtuse polygons identified in Theorem 1. Choose an infinite strip S of minimal width bounded by parallel perforation lines l and m, and "accordion-fold" T onto S, i.e., fold along l then along m so that S has four leaves configured as a "w," then fold again along l and again along m so that S has eight "zig-zag" leaves, and continue in this manner indefinitely.

Choose a stamp P in S. If P is a rectangle, accordion-fold S onto P. If P is a right triangle, P together with some subset of its images tessellate a rectangle R of minimal area contained in S, so accordion-fold S onto R, then fold R onto P. If P is an equilateral triangle, two of its edges lie in the interior of S. Label these edges a and b, then its third edge c is contained in the boundary of S. Let S_1 and S_2 be the subsets of $S - P$ bounded by l, m, and a, and by l, m, and b, respectively. Fold S_1 along a, then along c, then along b, and continue in this manner indefinitely to form an infinite "spiral"; similarly, fold S_2 along b, then along c, then along a, and continue indefinitely.

On the other hand, if T is generated by an obtuse polygon G, it has a 3-center C shared by three copies of G. Since the interior angle of G at C is bisected by a line of symmetry l, which contains an edge of some copy of G, folding along l creases the stamp G. Thus T is not suitable for stamp folding puzzles, and we have established Frederickson's conjecture:

THEOREM 2. *The edge tessellations suitable for stamp folding puzzles are generated by the four non-obtuse polygons indicated in Theorem 1.*

To summarize, we have proved that a polygon generating an edge tessellation is one of the following eight types: a rectangle; an equilateral, 60-right, isosceles right, or 120-isosceles triangle; a 120-rhombus; a 60-90-120 kite; or a regular hexagon. Of these, the four non-obtuse polygons generate tessellations suitable for stamp folding puzzles. Our proof of suitability, which establishes Frederickson's Conjecture, exhibits explicit algorithms for folding each suitable edge tessellation into a packet of single stamps.

Acknowledgment We wish to thank Andrew Hall and Joshua York for enthusiastically sharing their creative ideas in the early stages of this project, Natalie Frank for sharing her thoughts on stamp folding algorithms, and Andrew Baxter, Deirdre Smeltzer, Jim Stasheff, and Doris Schattschneider for offering many helpful editorial suggestions.

REFERENCES

1. A. Baxter, J. Eskreis-Winkler, and E. McCarthy, Periodic Billiard Paths in Edge-Tessellating Polygons, unpublished manuscript.
2. A. Baxter and R. Umble, Periodic Orbits for Billiards on an Equilateral Triangle, *Amer. Math. Monthly*, **115** (**6**) (2008) 479–491.
3. G. N. Frederickson, *Piano-Hinged Dissections: Time to Fold!* A. K. Peters, Ltd., Wellesley, MA, 2006.
4. B. Grünbaum and G. Shephard, *Tilings and Patterns*, W. H. Freeman and Company, New York, 1986.
5. G. E. Martin, *Transformation Geometry: An Introduction to Symmetry*, Springer-Verlag, New York, 1982.

Summary An edge tessellation is a tiling of the plane generated by reflecting a polygon in its edges. In this article we prove that a polygon generating an edge tessellation is one the following eight types: a rectangle; an equilateral, 60-right, isosceles right, or 120-isosceles triangle; a 120-rhombus; a 60-90-120 kite; or a regular hexagon. A stamp folding puzzle is a paper folding problem constrained to the perforations on a sheet of postage stamps. Such sheets necessarily embed in an edge tessellation. On page 143 of his book *Piano-Hinged Dissections: Time to Fold!*, G. Frederickson poses the following conjecture: "Although triangular stamps have come in a variety of different triangular shapes, only three shapes seem suitable for [stamp] folding puzzles: equilateral, isosceles right triangles, and 60°-right triangles." We prove that the four non-obtuse polygons mentioned above generate edge tessellations suitable for stamp folding puzzles. Our proof of suitability, which establishes Frederickson's Conjecture, exhibits explicit algorithms for folding each suitable edge tessellation into a packet of single stamps.

Folding Noneuclidean Strips of Paper

NIKOLAI A. KRYLOV
Siena College
Loudonville, NY 12211
nkrylov@siena.edu

EDWIN L. ROGERS
Siena College
Loudonville, NY 12211
rogers@siena.edu

Folding in the Euclidean plane

In the second edition of his book, *The Art and Craft of Problem Solving* [**8**], Paul Zeitz describes a procedure whose outcome is well known [**6, 3**].

Starting with a long strip of paper, hold the top left corner and fold it down so that the corner is now below the bottom of the strip. Unfold. Now, grab the right end of the strip and fold down so that the top side of the strip is along the crease just created. Unfold again. You now have two creases. Continue by grasping the right end and folding up so that the bottom side of the strip is along the most recently created crease. Unfold again and repeat until you have exhausted your strip (see FIGURE 1).

Math. Mag. **84** (2011) 289–294. doi:10.4169/math.mag.84.4.289.

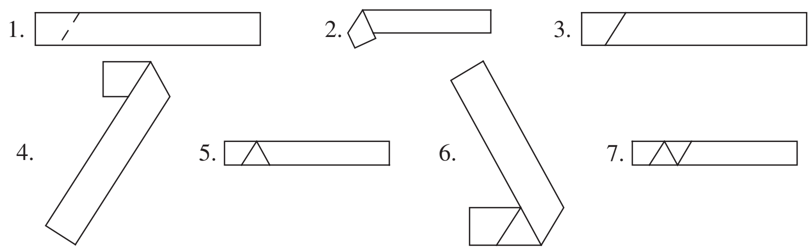


Figure 1 Folding a strip of paper

If you haven't tried this before, why not try it now? You will find that, after several iterations, the triangles created look more equilateral, and this appears to be independent of the angle of the initial crease. We give a simple derivation of this result that we will use again later. In FIGURE 2, we model the strip and initial crease with two parallel lines ℓ and m and transversal t. At each step, the procedure bisects the supplement of the angle created by the preceding step. Specifically, at step n, the angles satisfy $2\alpha_{n+1} + \beta_n = \pi$ and $2\beta_n + \alpha_n = \pi$, so it is easy to see that

$$\alpha_{n+1} = \frac{1}{4}\alpha_n + \frac{\pi}{4}.$$

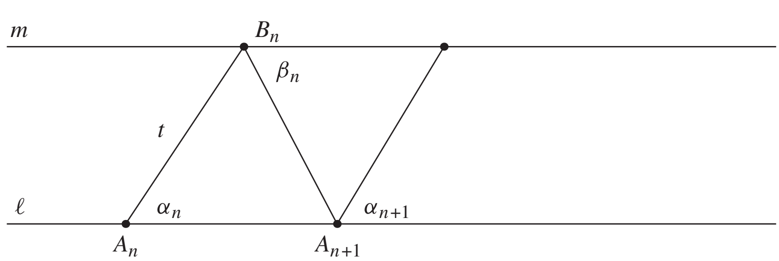


Figure 2 Folding in the Euclidean plane

This equation can be rewritten as $\left(\alpha_{n+1} - \frac{\pi}{3}\right) = \frac{1}{4}\left(\alpha_n - \frac{\pi}{3}\right)$, which implies that α_n converges to $\pi/3$. In fact, the deviation from $\pi/3$ at each stage is quartered, accounting for the rapid appearance of the equilateral triangle. An identical analysis shows that the sequence of angles, β_n, also converges to $\pi/3$.

We wondered if this property is unique to the Euclidean plane. Might similar results hold in the hyperbolic plane and and on the sphere? Of course, in hyperbolic geometry there are several types of parallel lines and on the sphere there are no parallel lines at all, so some modifications will need to be made. In the next section we explore the hyperbolic case, and we then consider the procedure on the sphere. We finish with some possible generalizations.

Folding a hyperbolic strip

In hyperbolic geometry, given a line and a point not on the given line, there are infinitely many lines through the given point parallel to the given line. This is sometimes referred to as the Hyperbolic Parallel Postulate (HPP). The basic properties of hyperbolic geometry can be found, for example, in the texts by Anderson [**1**] or Henle [**2**]. We mention here a few facts which will suffice for our needs.

First, one can distinguish between two types of parallelism. Two lines ℓ and m are said to be *asymptotically parallel* if the distance between them approaches zero as one moves in one direction and is unbounded as one moves in the opposite direction. By contrast, two lines are called *ultraparallel* or *hyperparallel* if they share a common perpendicular and the distance between them is unbounded as one moves away from the common perpendicular in either direction. For asymptotically parallel lines, a variable *angle of parallelism* measures the degree to which the parallelism is "noneuclidean" at any particular point. Let ℓ and m be asymptotically parallel, and let t be a transversal line which intersects ℓ and m at points A and B respectively, and is perpendicular to ℓ. Then on the side of t where the distance between ℓ and m decreases, the angle φ determined by t and m is strictly less than $\pi/2$. This is the angle of parallelism of ℓ at B (see FIGURE 3). It is not constant, but depends on the distance d (the length of the segment AB of t). If the unit of distance is chosen as the distance corresponding to the angle of parallelism $2\arctan(e^{-1})$, then Lobatchevsky's formula

$$e^{-d} = \tan(\varphi/2)$$

holds. As we move along m in the direction of decreasing distance, called the *direction of parallelism*, φ increases and approaches $\pi/2$. Thus, as we fold along two asymptotically parallel lines in the direction of parallelism, the process becomes more Euclidean, which suggests that the limiting triangle is again equilateral.

Perhaps the most striking departure from Euclidean geometry, and directly attributable to the HPP, is the fact that the sum of the interior angles of a triangle is no longer equal to π. For any triangle, the angle sum is always less than π and, as will be seen, this angle sum can take on any value in $[0, \pi)$. The difference between the angle sum of a given triangle and π defines its area. To be more precise, if we have a triangle $\triangle ABC$ with interior angles α, β, and γ, then with the unit chosen as above

$$\text{area} = \mu(ABC) = \pi - (\alpha + \beta + \gamma).$$

As a consequence, we see that triangle area is bounded. If we allow our triangles to have a pair of sides that are asymptotically parallel, a so-called asymptotic triangle, then the corresponding interior angle has measure 0. For example, in FIGURE 3 the points A and B are the vertices of a right asymptotic triangle with area $\pi/2 - \varphi$, a result that we apply below. If all pairs of adjacent sides of a triangle are asymptotically parallel, the triangle will has area π.

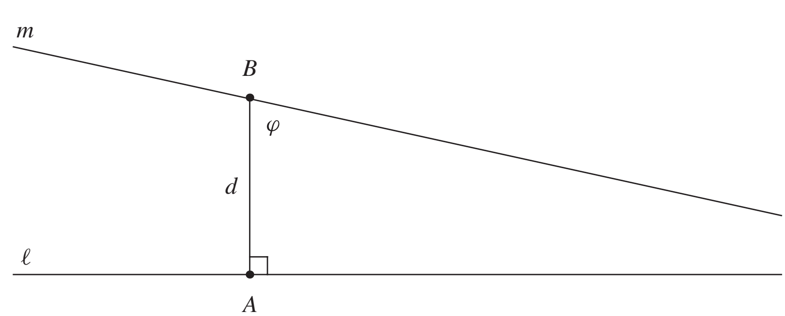


Figure 3 The angle of parallelism of ℓ at B

For ultraparallel lines ℓ and m, the folding procedure on a strip with these lines as edges breaks down after a finite number of steps. Moving along m away from the common perpendicular p, the angle of parallelism of ℓ at each B on m decreases towards zero as the distance between ℓ and m increases. Therefore, lines through B which intersect ℓ will make smaller and smaller angles with BA, A being the foot of the perpendicular from B to ℓ. Let P and Q denote the intersections of p with m and ℓ. The quadrilateral determined by P, Q, A, and B, known as a *Lambert quadrilateral*, has an acute angle at B, so its supplement is obtuse. If our procedure always produced a transversal to ℓ and m, then the bisection angle would have to approach zero. But from the comment above, this angle is bounded below by $\pi/4$. Therefore, the procedure must terminate.

So, to continue, we consider a strip bounded by pairs of lines that are asymptotically parallel. Again, denote the edges of the strip by ℓ and m and assume they are asymptotically parallel, and let t be an arbitrary transversal. Labeling the points of intersection, let $A_1 \in \ell$ and $B_1 \in m$, and apply the iterative folding procedure described above, folding between the lines in the direction of parallelism. First observe that the sequence of points, A_n and B_n, created on ℓ and m is monotone. They are unbounded as well. One can argue this as follows. Suppose to the contrary that each sequence of points is bounded and set $A = \lim_{n\to\infty} A_n$ and $B = \lim_{n\to\infty} B_n$. For any triple of points A_n, B_n, A_{n+1}, $\angle A_{n+1}A_nB_n$ is acute (with the possible exception of the initial triple) and the foot of the perpendicular, F_n from B_n to ℓ must lie between A_n and A_{n+1}. This follows since $\angle A_nB_nF_n$ must have measure less than $\angle F_nB_nB_{n+1}$, the angle of parallelism. If these measures were equal, the line determined by A_n, B_n would be asymptotically parallel to ℓ in the other direction, contradicting the fact that A_n lies on ℓ. Letting $F = \lim_{n\to\infty} F_n$, it follows that $A = F$ and the segment BF is perpendicular to ℓ at A. Similarly, the foot of the perpendicular, G_n from A_n to m will converge to B, and the segment, AG will be perpendicular to m at B, giving the two asymptotically parallel lines ℓ and m a common perpendicular, a contradiction.

Referring to FIGURE 4, we can obtain a difference equation for angles as follows. Noting that $\mu(A_nB_nA_{n+1}) = \pi - (\alpha_n + \beta_n + \pi - 2\alpha_{n+1})$ and $\mu(A_{n+1}B_n) = \pi - (2\alpha_{n+1} + \beta_n)$ (where μ is the area function defined above and $\mu(A_{n+1}B_n)$ is the area of the asymptotic triangle with vertices A_{n+1} and B_n), solving for α_{n+1} yields

$$\alpha_{n+1} = \frac{1}{4}\alpha_n + \frac{\pi}{4} + \frac{\mu(A_nB_nA_{n+1}) - \mu(A_{n+1}B_n)}{4}. \tag{1}$$

Now note that $\triangle A_nB_nA_{n+1}$ and $\triangle A_{n+1}B_n$ are both contained in the asymptotic right triangle $\triangle A_nG_n$, where G_n is the foot of the perpendicular from A_n to m. As noted above, this triangle has area $\pi/2 - \varphi$, where φ is the angle of parallelism at A_n. As folding proceeds in the direction of parallelism, φ approaches $\pi/2$ and the area of $\triangle A_nG_n$ approaches zero. Therefore, the areas $\mu(A_nB_nA_{n+1})$ and $\mu(A_{n+1}B_n)$ must approach zero as well.

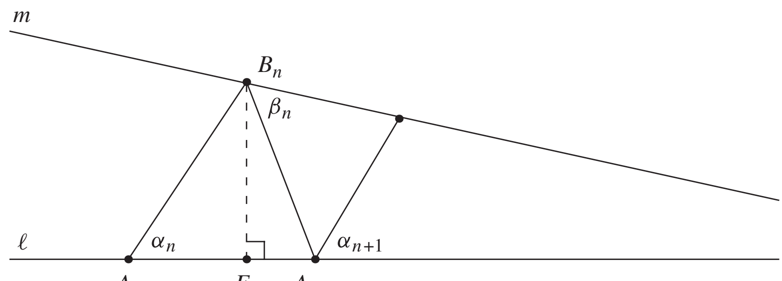


Figure 4 Folding along asymptotic parallels

Now rewrite as before to conclude that α_n approaches $\pi/3$. The same analysis can be performed from the point of view of β_n, and the conclusion remains the same. Thus, the limiting triangle must be equilateral.

Folding a strip on the sphere

We now imagine our strip lying on a sphere. In the geometry of the sphere there are no parallel lines and two points may not determine a unique line, although in what follows we restrict ourselves to regions where this is true. As before, we begin with two lines ℓ and m with an arbitrary transversal t, but this time the lines intersect at some point P with angle ϕ. We are interested in folding in the direction of the point of intersection. Triangle area on the sphere is akin to its hyperbolic counterpart. The natural unit for measuring distance in this case is the radius, and in particular $r = 1$ suits our purpose here. With this choice, the spherical triangle $\triangle ABC$ with interior angles α, β, and γ has area given by

$$\text{area} = \mu(ABC) = \alpha + \beta + \gamma - \pi.$$

As an illustration, consider the triangle $\triangle ABC$ shown in FIGURE 5, which is obtained by cutting the sphere into 8 identical triangles; four on the upper hemisphere and four on the lower. Each triangle has three right angles for its interior angles. Since area of such a sphere is 4π, the area formula above clearly holds.

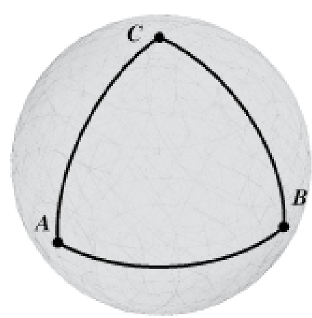


Figure 5 A spherical triangle

Using FIGURE 6 as a guide, we find as before that

$$\alpha_{n+1} = \frac{1}{4}\alpha_n + \frac{\pi - \phi}{4} - \frac{\mu(A_n B_n A_{n+1}) - \mu(A_{n+1} B_n P)}{4} \tag{2}$$

As we fold towards the point of intersection, A_n and B_n both approach P. It is not then difficult to see that both areas $\mu(A_n B_n A_{n+1})$ and $\mu(A_{n+1} B_n P)$ approaches zero. On subtracting $\frac{\pi-\phi}{3}$ from both sides of (2) and again rewriting, α_n is now seen to approach $\frac{\pi-\phi}{3}$. In like manner, β_n will approach the same value, and we obtain in this case an isosceles triangle in the limit.

Further investigations

Based on our experience with the sphere, one can look at folding between lines in general. It is not difficult to show that intersecting lines in the Euclidean and hyperbolic planes provide results similar to what was discovered on the sphere. As a natural

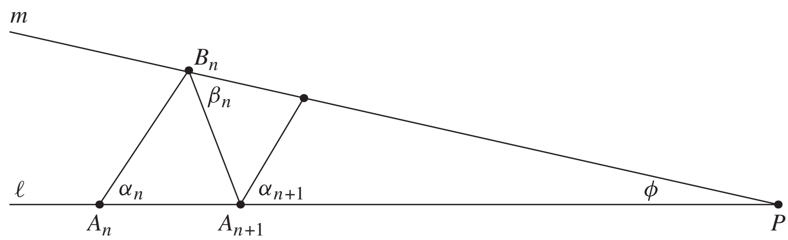


Figure 6 Folding in the sphere

generalization of the procedure, one could fold with ratio q, where the folding angles at each stage are in ratio q, and more generally fold with alternating and distinct ratios at each step. One could also weaken the assumptions on the bounding lines ℓ and m. For example, ℓ and m may be asymptotic plane curves or plane curves asymptotic to lines that intersect transversally. The authors have investigated some of these questions, along with generalizations to more general surfaces in [**4**], [**5**].

In closing, we note some recent history regarding questions of this type in the work of J. Pedersen [**6**]; P. Hilton, D. Holton, and J. Pedersen [**3**]; and more recently B. Polster [**7**]. These authors demonstrate methods for constructing polygons and polyhedra from strips of paper. In each of these works, an essential feature in the construction is that the "final" angle is independent of the initial angle. This is shown to have practical value when one would like to construct an arbitrary rational angle (that is, angles of the form $a\pi/n$ with a and n relatively prime) by simple folding.

REFERENCES

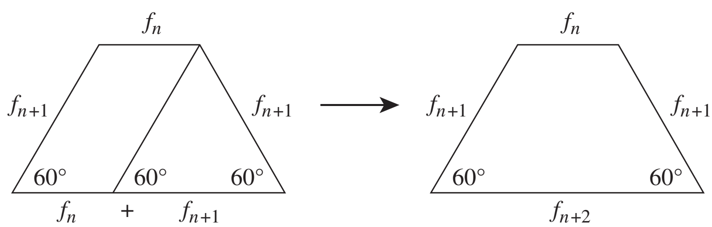
1. James W. Anderson, *Hyperbolic Geometry*, Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, 2nd ed., 2005.
2. Michael Henle, *Modern Geometries: Non-Euclidean, Projective, and Discrete*, Prentice Hall, Inc., Upper Saddle River, NJ, 2nd ed., 2001.
3. Peter Hilton, Derek Holton, and Jean Pedersen, *Mathematical Reflections. In a Room With Many Mirrors*, Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1997.
4. Nikolai Krylov and Edwin Rogers, Dynamics of folds in the plane, *College Math. J* **42** (2011), 111–117.
5. Nikolai Krylov and Edwin Rogers, Angle contraction between geodesics, *Journal of Dynamical Systems and Geometric Theories* **8** (2010) 1–9.
6. Jean Pedersen, Combinatorics, group theory and geometric models, *Cahiers Topologie Géom. Différentielle* **22** (1981) 407–428.
7. Burkard Polster, Variations on a theme in paper folding, *Amer. Math. Monthly* **111** (2004) 39–47. http://dx.doi.org/10.2307/4145014
8. Paul Zeitz, *The Art and Craft of Problem Solving*, John Wiley & Sons, Inc., New York, 2nd ed., 2007.

Summary We have considered the process of inscribing triangles between parallel lines in the Euclidean plane using an iterative process. It is a familiar result that the limiting triangle is equilateral. In this Note, we demonstrate that the same result holds when the geometry is hyperbolic and the lines are asymptotically parallel. We further extend the process to the sphere, where it is found that the limiting triangle is isosceles.

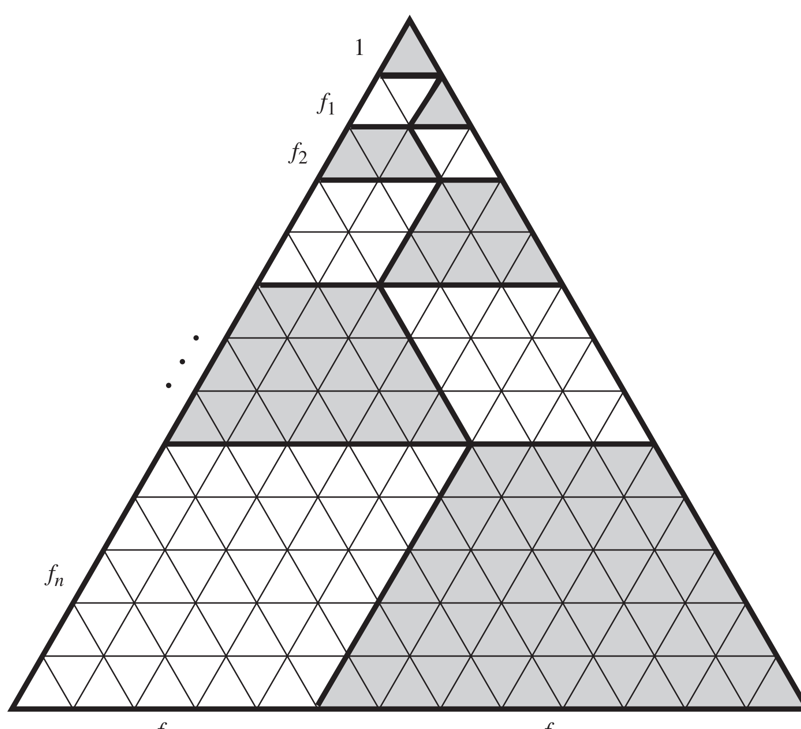
Proof Without Words: Fibonacci Trapezoids

HANS R. WALSER
Mathematisches Institut Uni Basel
Rheinsprung 21, CH-4051 Basel
Switzerland
hwalser@bluewin.ch

I. Fibonacci Recursion: $f_n + f_{n+1} = f_{n+2}$



II. Identity: $1 + \sum_{k=1}^{n} f_k = f_{n+2}$



Math. Mag. **84** (2011) 295. doi:10.4169/math.mag.84.4.295.

PROBLEMS

PROPOSALS

To be considered for publication, solutions should be received by March 1, 2012.

1876. *Proposed by Roman Wituła, Edyta Hetmaniok, and Damian Słota, Institute of Mathematics, Silesian University of Technology, Gliwice, Poland.*

Prove that the following equality holds for $x, y \in \mathbb{C}$ and n a positive integer.

$$x^{2n} - x^n y^n + y^{2n} = \prod_{\substack{1 \le k < 3n \\ \gcd(k,6)=1}} \left(x^2 - 2\cos\left(\frac{k\pi}{3n}\right) xy + y^2 \right).$$

1877. *Proposed by Daniel Edelman, Mason–Rice Elementary School, Newton Centre, MA, and Alan Edelman, MIT, Cambridge, MA.*

For each $n \times n$ permutation matrix M, consider the graph G_M where the vertices are the zero entries and two vertices are adjacent if their corresponding entries in the matrix are adjacent horizontally or vertically. We say that M *disconnects its zeros*, if G_M is disconnected. For example, M_1 has the bottom left zero disconnected, while M_2 does not disconnect its zeros.

$$M_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Find a formula for the number of $n \times n$ permutation matrices that disconnect their zeros. Also find an asymptotic formula as $n \to \infty$ for the fraction of the $n!$ permutation matrices that disconnect their zeros.

Math. Mag. **84** (2011) 296–303. `doi:10.4169/math.mag.84.4.296`.

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

Solutions should be written in a style appropriate for this MAGAZINE.

Solutions and new proposals should be mailed to Bernardo M. Ábrego, Problems Editor, Department of Mathematics, California State University, Northridge, 18111 Nordhoff St, Northridge, CA 91330-8313, or mailed electronically (ideally as a LaTeX or pdf file) to `mathmagproblems@csun.edu`. All communications, written or electronic, should include **on each page** the reader's name, full address, and an e-mail address and/or FAX number.

1878. *Proposed by Pantelimon George Popescu, Politechnica University, Bucharest, Romania and José Luis Díaz-Barrero, Polytechnical University of Catalonia, Barcelona, Spain.*

Let $\alpha \geq -1$ and $\beta \geq 1$ be real numbers. Let $\{b_k : 1 \leq k \leq n\}$ be a set of real numbers in the interval $(0, 1]$. Prove that

$$\frac{A-\alpha}{A+\beta} \geq \frac{1}{n}\sum_{k=1}^{n}\frac{b_k-\alpha}{b_k+\beta} \geq \frac{G-\alpha}{G+\beta},$$

where A and G are the arithmetic and geometric mean, respectively, of the set $\{b_k : 1 \leq k \leq n\}$.

1879. *Proposed by Wong Fook Sung, Temasek Polytechnic, Singapore.*

Let m and n be positive integers such that $m < n$ and let a and b be positive real numbers. Evaluate

$$\int_0^\infty \frac{x^{2(n-m)}(x^2-1)^{2m}}{ax^{2n}+b(x^2-1)^{2n}}\,dx.$$

1880. *Proposed by Richard Stephens, Department of Mathematics, Columbus State University, Columbus, GA.*

Let X be a positive continuous random variable and for any $\alpha > 0$, let Y_α be the random variable defined by $Y_\alpha = n$ for $n = 1, 2, 3, \ldots$ if and only if $(n-1)\alpha < X < n\alpha$. Prove that X has an exponential distribution if and only if Y_α has a geometric distribution for every $\alpha > 0$.

Quickies

Answers to the Quickies are on page 302.

Q1013. *Proposed by Michel Bataille, Rouen, France.*

Let a, b, c be positive real numbers such that $abc \geq 1$. Prove that

(a) $4(a^4+b^4+c^4) \geq 3\sqrt[3]{(a+b)^2(b+c)^2(c+a)^2}$, and
(b) $(2a^4+b^4)(2b^4+c^4)(2c^4+a^4) \geq (2a+b)(2b+c)(2c+a)$.

Q1014. *Proposed by Ovidiu Furdui, Campia Turzii, Cluj, Romania.*

Find the limit

$$\lim_{n\to\infty}\left(\int_0^1 (1+x^n)^n\,dx\right)^{1/n}.$$

Solutions

A recursive sequence that respects the primes **October 2010**

1851. *Proposed by Éric Pité, Paris, France.*

Let a be an arbitrary integer. Consider the recursive sequence of integers defined by $u_0 = 4$, $u_1 = 0$, $u_2 = 2$, $u_3 = 3$, and $u_{n+4} = u_{n+2} + u_{n+1} + a \cdot u_n$ for every integer $n \geq 0$. Prove that p divides u_p for every prime p.

Solution by John L. Simons, University of Groningen, the Netherlands.

Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ be the roots of the characteristic polynomial $f(x) = \prod_{i=1}^{i=4}(x - \alpha_i) = x^4 - x^2 - x - a$. By comparison of coefficients we find that

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0,$$
$$\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_1\alpha_4 + \alpha_2\alpha_3 + \alpha_2\alpha_4 + \alpha_3\alpha_4 = -1,$$
$$\alpha_1\alpha_2\alpha_3 + \alpha_1\alpha_2\alpha_4 + \alpha_1\alpha_3\alpha_4 + \alpha_2\alpha_3\alpha_4 = 1, \text{ and}$$
$$\alpha_1\alpha_2\alpha_3\alpha_4 = -a.$$

Consider the potential solution $u_n = \sum_{i=1}^{i=4} \alpha_i^n$. It is easily seen that this solution satisfies the recursive equation $u_{n+4} = u_{n+2} + u_{n+1} + a \cdot u_n$. For the initial values we find that

$$
\begin{aligned}
u_0 &= 4,\\
u_1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0,\\
u_2 &= \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)^2 - 2(\alpha_1\alpha_2 + \ldots + \alpha_3\alpha_4) = 2, \text{ and}\\
u_3 &= \alpha_1^3 + \alpha_2^3 + \alpha_3^3 + \alpha_4^3\\
&= (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)^3 - 3(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)\\
&\quad (\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_1\alpha_4 + \alpha_2\alpha_3 + \alpha_2\alpha_4 + \alpha_3\alpha_4)\\
&\quad\quad + 3(\alpha_1\alpha_2\alpha_3 + \alpha_1\alpha_2\alpha_4 + \alpha_1\alpha_3\alpha_4 + \alpha_2\alpha_3\alpha_4) = 3.
\end{aligned}
$$

Consequently $u_n = \sum_{i=1}^{i=4} \alpha_i^n$ is the general solution of the recursive equation with the given initial values. For p prime, all binomial coefficients $\binom{p}{j}$ with $0 < j < p$ contain a factor p. Thus

$$u_p = \alpha_1^p + \alpha_2^p + \alpha_3^p + \alpha_4^p = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)^p - p \cdot g(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$$

where $g(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is an integral symmetric polynomial in the roots α_i. A basic theorem in the theory of symmetric polynomials states that such a symmetric polynomial in the roots α_i can be integrally expressed in the coefficients of the characteristic polynomial. From this it follows that for p prime,

$$u_p \equiv (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)^p \equiv 0 \pmod{p}.$$

Editor's Note. Several solvers pointed out that a more general theorem by L. E. Dickson appeared in his solution to Problem 151 in *Amer. Math. Monthly* **15** (1908), 209. Other approaches to the problem are mentioned in a note by Gregory Minton, this *Magazine* **84** (2011), 33.

Also solved by George Apostolopulous (Greece), Michel Bataille (France), John Christopher, Con Amore Problem Group (Denmark), Amanda Goodrick, G.R.A.20 Problem Solving Group (Italy), Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Germany), Peter McPolin (Ireland), Rituraj Nandan, José Heber Nieto (Venezuela), Nicholas C. Singer, Marian Tetiva (Romania), Li Zhou, and the proposer. There was one incomplete submission.

Integrating the square of the derivative **October 2010**

1852. *Proposed by Radu Gologan, Institute of Mathematics "Simion Stoilow" of the Romanian Academy, Bucharest, Romania; and Cezar Lupu (student), University of Bucharest, Bucharest, Romania.*

Let $f : [0, 1] \to \mathbb{R}$ be a differentiable function with a continuous derivative such that $f(0) = f(1) = -\frac{1}{6}$. Prove that

$$\int_0^1 (f'(x))^2\, dx \geq 2\int_0^1 f(x)\, dx + \frac{1}{4}.$$

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

First, we see that

$$\begin{aligned} 0 &\leq \int_0^1 \left(f'(x) + x - \frac{1}{2}\right)^2 dx \\ &= \int_0^1 (f'(x))^2\, dx + \int_0^1 (2x-1) f'(x)\, dx + \int_0^1 \left(x - \frac{1}{2}\right)^2 dx, \end{aligned}$$

where the last integral equals $\frac{1}{12}$. Integration by parts and the hypothesis that $f(0) = f(1) = -\frac{1}{6}$ give

$$\int_0^1 (2x-1) f'(x)\, dx = (2x-1) f(x)\Big|_{x=0}^{x=1} - 2\int_0^1 f(x)\, dx = -\frac{1}{3} - 2\int_0^1 f(x)\, dx.$$

Substituting this into the initial inequality gives

$$0 \leq \int_0^1 (f'(x))^2\, dx - 2\int_0^1 f(x)\, dx - \frac{1}{4},$$

which completes the proof.

Editor's Note. A considerable number of solutions made use of the Cauchy–Schwarz Inequality. Several solvers pointed out that the inequality turns into an equality if and only if $f(x) = \frac{1}{2}x(1-x) - \frac{1}{6}$, $x \in [0, 1]$. William R. Green noted that with the same method one can generalize the problem to: "For f continuously differentiable on $[0, a]$ $(a > 0)$ and $f(a) = f(0) = b$ we have that

$$\int_0^a (f'(x))^2\, dx \geq 2\int_0^a f(x)\, dx - \left(2ab + \frac{a^3}{12}\right)."$$

Also solved by George Apostolopulous (Greece), Michel Bataille (France), Andrew Gibson and Cecil Rousseau, Michael W. Botsko, Bill Cowieson, Robert L. Doucette, G.R.A.20 Problem Solving Group (Italy), William R. Green, Eugene A. Herman, Enkel Hysnelaj (Australia) and Elton Bojaxhiu (Germany), Jeff Ibbotson, Omran Kouba (Syria), Kee-Wai Lau (China), Peter McPolin (Ireland), Elio Miranda (Argentina), Ariel Norambuena (Chile), Hyeonjune Park (South Korea), Paolo Perfetti (Italy), David S. Ross, Nicholas C. Singer, D. V. Thong (Vietnam), Tiberiu Trif (Romania), Michael Vowe (Switzerland), and the proposers. There was one incorrect submission.

Nearly continuous functions characterized **October 2010**

1853. *Proposed by Michael W. Botsko, Saint Vincent College, Latrobe, PA.*

A function f is continuous nearly everywhere if it is continuous on its domain except for a countable set. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function.

(a) Prove that if f is continuous nearly everywhere, then for every open set $G \subseteq \mathbb{R}$ there are an open set O and a countable set C such that $f^{-1}(G) = O \cup C$.

(b) Is the converse of part (a) true? Prove or disprove.

Solution by Bill Cowieson, Fullerton College.

(a) We prove the statement for $f : X \to Y$, where X and Y are any topological spaces. Let $C(f) \subset X$ be the set of continuity points for f and let $D(f) = X \backslash C(f)$ be the (countable) set of discontinuity points for f. For any open subset $G \subset Y$ and any point $x \in f^{-1}(G) \cap C(f)$, the openness of G and the continuity of f at x guarantee an open neighborhood U_x of x with $f(U_x) \subset G$. It follows that $U_x \subset f^{-1}(G)$. Define

$$O = \bigcup_{x \in f^{-1}(G) \cap C(f)} U_x \quad \text{and} \quad D = f^{-1}(G) \backslash O.$$

The set D is countable since it consists solely of discontinuity points, O is an open subset of $f^{-1}(G)$, and $O \cup D = f^{-1}(G)$.

(b) We prove the converse for $f : X \to Y$, where X is any topological space and Y is second countable. Let $\mathcal{B} = \{B_1, B_2, \dots\}$ be a countable base for the topology on Y. By assumption, for each B_i there exists an open set $O_i \subset X$ and a countable set $D_i \subset X$ with $f^{-1}(B_i) = O_i \cup D_i$. The set $D = \bigcup_i D_i$ is countable, and we claim that f is continuous on its complement, $X \backslash D$. To see this, consider any $x \notin D$ and any open neighborhood G of $f(x)$. Since $\mathcal{B}$ is a base for Y, there exists a B_i with $f(x) \in B_i \subset G$, so $x \in f^{-1}(f(x)) \subset f^{-1}(B_i) = O_i \cup D_i$. In fact, $x \in O_i$ since we chose $x \notin D$, so we have shown that for any neighborhood G of $f(x)$, there is a neighborhood $O = O_i$ of x with $f(O) \subset G$, that is, f is continuous at x. Since $x \notin D$ is arbitrary, this proves that f is continuous nearly everywhere on X.

Note. The second countability assumption is necessary, as this next example shows. Let $X = (\mathbb{R}, \mathcal{T})$ be the real numbers with the usual topology generated by intervals of the form (a, b), let $Y = (\mathbb{R}, \mathcal{L})$ be the real numbers with the "lower limit" topology generated by intervals of the form $[a, b)$, and let $f : \mathbb{R} \to \mathbb{R}$ be the identity. The function f satisfies the assumptions since any $G \in \mathcal{L}$ can be written as a countable union $G = \cup_i [a_i, b_i)$ for some real numbers $\{a_i\}$ and $\{b_i\}$, so

$$f^{-1}(G) = G = \bigcup_i [a_i, b_i) = \bigcup_i (a_i, b_i) \cup \bigcup_i \{a_i\} = O \cup D,$$

where $O \in \mathcal{T}$ and D is countable. However, f is nowhere continuous since for any $x \in X$, $[x, x+1)$ is an $\mathcal{L}$-neighborhood of $f(x) = x \in Y$, but $f^{-1}([x, x+1)) = [x, x+1)$ is not a $\mathcal{T}$-neighborhood of $x \in X$.

Also solved by Paul Budney, Bruce S. Burdick, Eugene A. Herman, Bianca-Teodora Iordache (Romania), Mathramz Problem Solving Group, José H. Nieto (Venezuela), Northwestern University Math Problem Solving Group, Paolo Perfetti (Italy), Texas State University Problem Solvers Group, and the proposer.

Subsets that make no difference *d*. October 2010

1854. *Proposed by Marian Tetiva, National College "Gheorghe Roşca Codreanu", Bârlad, Romania.*

Let n and d be nonnegative integers. Find the number of all subsets of $\{1, 2, \dots, n\}$ which do not contain two numbers whose difference is d. (Subsets with at most one element satisfy the condition by vacuity.)

Solution by José Heber Nieto, Universidad del Zulia, Maracaibo, Venezuela.

Let q and r be nonnegative integers such that $n = qd + r$ and $0 \le r < d$. Then the answer is $F_{q+3}^r F_{q+2}^{d-r}$, where F_k denotes the Fibonacci number defined by $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n > 1$.

Given n integers $a_1 < a_2 < \ldots < a_n$, if we call a_i and a_{i+1} neighbors, it is well known that the number x_n of subsets of $A = \{a_1, a_2, \ldots, a_n\}$ without neighbors is F_{n+2}. Indeed, for $n > 1$ these subsets may be partitioned into those which contain a_1, which are x_{n-2}, and those which do not, which are x_{n-1}. Thus $x_n = x_{n-1} + x_{n-2}$, the same recurrence satisfied by the Fibonacci numbers. Since $x_0 = 1 = F_2$ and $x_1 = 2 = F_3$, it follows that $x_n = F_{n+2}$. Now let $I_n = \{1, 2, \ldots, n\}$ and, for $k = 1, 2, \ldots, d$, let $C_k = \{x \in I_n : x \equiv k \pmod d\}$. Thus I_n is the disjoint union of $C_1, \ldots, C_d$. Since two integers in I_n whose difference is d must be in the same C_i, it is clear that any subset $S \subseteq I_n$ which do not contain two numbers whose difference is d, is the union of d subsets $S_i \subseteq C_i$, for $k = 1, 2, \ldots, d$, such that each S_i is a subset of C_i without neighbors. Since $C_1, \ldots, C_r$ contain $q + 1$ elements each, while $C_{r+1}, C_{r+2}, \ldots, C_d$, contain q elements each, it follows that the answer is $F_{q+3}^{r} F_{q+2}^{d-r}$.

Note. The answer above still holds if $d > n$. In this case, as well as when $d = 0$, all 2^n subsets of I_n satisfy the condition.

Also solved by Michael Andreoli; George Apostolopoulos (Greece); Elton Bojaxhiu (Germany) and Enkel Hysnelaj (Australia); Daniele Degiorgi (Switzerland); Dmitry Fleischman; Marty Getz and Dixon Jones; Andrew Gibson, Cecil Rousseau, and Jonathan Hulgan; Carlo Pagano (Italy); Victor Y. Kutsenok; Mathramz Problem Solving Group; David Nacin; Northwestern University Math Problem Solving Group; Goran Ruzic (Canada); Joel Schlosberg; Nicholas C. Singer; Texas State University Problem Solvers Group; and the proposer.

Perpendicular analogies of Euler and Brocard lines **October 2010**

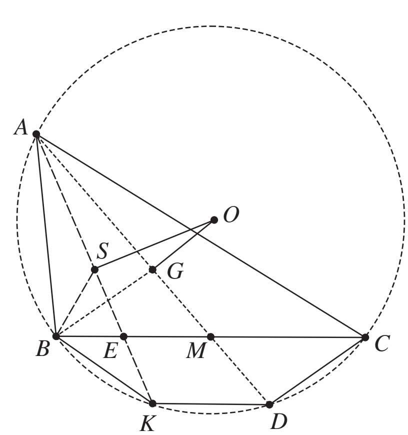
1855. *Proposed by Michael Goldenberg and Mark Kaplan, The Ingenuity Project, Baltimore Polytechnic Institute, Baltimore, MD.*

Prove that the Euler line of a triangle is perpendicular to one of the medians, if and only if, the Brocard line is perpendicular to the symmedian line from the same vertex as the median.

(The Brocard line passes through the circumcenter and the symmedian point of a triangle. The symmedian point is the point of concurrency of the symmedians and a symmedian through a vertex is the line symmetric to the median with respect to the angle bisector from the same vertex.)

Solution by Eunsoo Jee and Seonwoo Kim, Institute of Science Education for the Gifted and Talented, Yonsei University, Seoul, Korea.

Let O, G, and S be the circumcenter, the centroid, and the symmedian point of $\triangle ABC$, respectively. Let $\overline{AM}$ and $\overline{AE}$ be the median and the symmedian of $\triangle ABC$ by A, respectively. Finally, let D and K be the intersections of the circumcircle of $\triangle ABC$ with the lines AM and AE, respectively.



Note that the triangles ABM and AKC are similar because by symmedian definition $\angle BAM = \angle KAC$ and $\angle ABM = \angle ABC = \angle AKC$ because they subtend the same arc of circle. Thus $AB/AK = BM/KC$, and

$$AB \cdot KC = \frac{1}{2} BC \cdot AK. \tag{1}$$

Also, because $\overline{AK}$ is a symmedian, it follows that $\angle BAD = \angle KAC$, thus

$$\angle CBK = \angle CAK = \angle BAD = \angle BCD. \tag{2}$$

Last, because $\overline{BS}$ is a symmedian, it follows that

$$\angle ABS = \angle CBG. \tag{3}$$

Now we find a condition that is equivalent to both $\overline{OS} \perp \overline{AE}$ and $\overline{OG} \perp \overline{AM}$. On the one hand, $\overline{OS} \perp \overline{AE} \Leftrightarrow AS = SK$ (because $AO = OK$) $\Leftrightarrow BC \cdot AS = \frac{1}{2} BC \cdot AK \Leftrightarrow BC \cdot AS = AB \cdot KC$ (by (1)) $\Leftrightarrow \triangle ABS \sim \triangle CBK$ (because $\angle BAS = \angle BCK$) $\Leftrightarrow \angle ABS = \angle CBK \Leftrightarrow \angle CBG = \angle BCD$ (by 2) and (3)).

On the other hand, $\overline{OG} \perp \overline{AM} \Leftrightarrow AG = GD$ (because $AO = OD$) $\Leftrightarrow GM = MD$ (because $AG = 2GM$) $\Leftrightarrow \triangle BMG \cong \triangle CMD$ (because $BM = MC$ and $\angle BMG = \angle CMD$) $\Leftrightarrow \angle CBG = \angle BCD$.

Also solved by George Apostolopoulos (Greece), Michel Bataille (France), Elton Bojaxhiu (Germany) and Enkel Hysnelaj (Australia), Bruce S. Burdick, Chip Curtis, Michael Vowe (Switzerland), and the proposers.

Answers

Solutions to the Quickies from page 297.

A1013.
We start from Sophie Germain's famous identity: $4a^4 + b^4 = (a^2 + (a+b)^2)(a^2 + (a-b)^2)$ and deduce that

$$4a^4 + b^4 \geq a^2(a^2 + (a+b)^2). \tag{1}$$

(a) Since $a^2(a^2 + (a+b)^2) = a^4 + a^2(a+b)^2$, Inequality (1) gives $3a^4 + b^4 \geq a^2(a+b)^2$. Similarly, $3b^4 + c^4 \geq b^2(b+c)^2$ and $3c^4 + a^4 \geq c^2(c+a)^2$. Adding these inequalities gives

$$4(a^4 + b^4 + c^4) \geq a^2(a+b)^2 + b^2(b+c)^2 + c^2(c+a)^2.$$

The result follows by applying the Arithmetic Mean–Geometric Mean Inequality to the right-hand side, taking into account that $a^2b^2c^2 \geq 1$.

(b) Returning to Inequality (1) written as $4a^4 + b^4 \geq a^2(2a^2 + b^2 + 2ab)$, it follows that

$$2a^4 + b^4 \geq a^2b(2a + b).$$

Similarly, $2b^4 + c^4 \geq b^2c(2b + c)$ and $2c^4 + a^4 \geq c^2a(2c + a)$. Multiplying these inequalities gives

$$(2a^4 + b^4)(2b^4 + c^4)(2c^4 + a^4) \geq a^3b^3c^3(2a + b)(2b + c)(2c + a),$$

and the result follows since $abc \geq 1$.

A1014.

The limit equals 2. Let $x_n = \left(\int_0^1 (1+x^n)^n\,dx\right)^{1/n}$. First note that

$$\int_0^1 (1+x^n)^n\,dx \le \int_0^1 (1+1)^n\,dx = 2^n,$$

thus $x_n \le 2$. Second, the Arithmetic Mean–Geometric Mean Inequality implies that $1 + x^n \ge 2x^{n/2}$, which in turn gives

$$x_n \ge \left(\int_0^1 2^n x^{n^2/2}\,dx\right)^{1/n} = 2\left(\frac{2}{n^2+2}\right)^{1/n}.$$

Therefore

$$2 \ge \lim_{n\to\infty} x_n \ge \lim_{n\to\infty} 2\left(\frac{2}{n^2+2}\right)^{1/n} = 2.$$

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.

Frantz, Marc, and Annalisa Crannell, *Viewpoints: Mathematical Perspective and Fractal Geometry in Art*, Princeton University Press, 2011; xi + 232 pp, $45. ISBN 978-0-691-12592-3.

Have you ever wanted to team-teach with someone in your institution's Art Dept.? This book is a ready-made opportunity for a course involving mathematics that an artist colleague and art students would enjoy. The authors rightly describe this revolutionary book as "an undergraduate text in mathematics and art suitable for math-for-liberal-arts courses, mathematics courses for fine arts majors, and introductory art classes." It is "activity-based," with authentic "problems and activities of genuine interest and value to art students" that are also "real math problems." What art? Perspective (in various forms), vanishing points, anamorphic art, fractals. What mathematics? Space coordinates, similar triangles, affine transformations, exponents and logarithms (with sections reviewing their rules), and formulas to enter into spreadsheets (but no other software, e.g., for iterated function systems). There are exercises and activities, with selected answers. The book is beautiful in design and replete with figures and illustrations, including a dozen color plates and nine multi-page vignettes of contemporary artists and their work. For instructors, there is a detailed syllabus, timetable of how to use the book in connection with other sources (e.g., *Flatland*), suggested readings, and topics for writing assignments on the mathematics or art.

Livio, Mario, Why math works, *Scientific American* (August 2011) 80–83. http://www.scientificamerican.com/article.cfm?id=why-math-works .

Rowlett, Peter, et al., The unplanned impact of mathematics, *Nature* 475 (14 July 2011) 166–169.

Wigner, Eugene, The unreasonable effectiveness of mathematics in the natural sciences, *Communications in Pure and Applied Mathematics* 13 (1) (February 1960) 1–14.

Livio, an astrophysicist and the author of *Is God a Mathematician?* (2010), *The Equation That Couldn't Be Solved: How Mathematical Genius Discovered the Language of Symmetry* (2006), and *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number* (2003), ruminates on Eugene Wigner's marveling at the "unreasonable effectiveness of mathematics." Livio concludes that the explanation is that mathematics is both invented and discovered: "we [invent and] adopt mathematical tools," not just arbitrarily, but ones "that apply to our world"; and "scientists... select problems that are amenable to mathematical treatment." Meanwhile, Rowlett and colleagues tell seven short tales showing how theoretical mathematical developments—quaternions, Riemannian geometry, the E_8 lattice, Parrondo's paradox, the law of large numbers, topology, Fourier series—were followed by practical applications—computer graphics; cosmology; modems; pandemics; actuarial science; cellphones; and the quantum mechanics behind lasers, flat-screen TVs, and nuclear power.

Math. Mag. **84** (2011) 304–305. doi:10.4169/math.mag.84.4.304.

Devlin, Keith, *Mathematics Education for a New Era: Video Games as a Medium for Learning*, A K Peters, 2011; xiii + 203 pp. $29.95. ISBN 978-1-56881-431-5.

You, and most mathematics instructors, may soon be obsolete. Why? “[V]ideo games are going to play a major role in school mathematics education in the future.” Author Devlin mainly has middle school in mind, but aren’t students educated by games going to expect use of the the same medium in high school and college instruction? He grants that there are now no games that develop “mathematical thinking” (a topic that he delves into), only ones that focus on “math skills.” Devlin maintains that just as in learning to play an instrument, a skills-first approach is unsuitable. He cites well-known literature about children and adults who can solve arithmetic problems in their heads in the course of buying and selling but cannot do the equivalent when the problem is presented on paper or in a test-like setting. He focuses on the importance of “engagement” as the motivation for learning, and exhorts that exploration (play) precede instruction: Students can solve a problem “when they need it to do something that matters to them in the course of their everyday lives.” However, they are unlikely to need to slay a monster by solving a multiplication problem written on its chest. So just what problems do middle-schoolers (or others) have in their lives that can be addressed by mathematical thinking?? Devlin goes on to elaborate 11 “principles for an ideal learning environment” and enlarge on James Gee’s 36 principles for an educational game.

Baez, John C., and John Huerta, The strangest numbers in string theory, *Scientific American* (May 2011) 60–65.

———, The octonions, *Bulletin of the American Mathematical Society* 39 (2) (2002) 145–205. http://www.ams.org/journals/bull/2002-39-02/S0273-0979-01-00934-X/S0273-0979-01-00934-X.pdf .

———, and Helen Joyce, Ubiquitous octonions, *Plus* (Issue 33) (2005) http://plus.maths.org/content/ubiquitous-octonions .

Octonions, an 8-dimensional noncommutative and nonassociative “number system” (division algebra), were discovered in 1843 by John Graves (inspired by Hamilton) and rediscovered independently by Cayley in 1845. They turn out to be the kind of “numbers” used by matter and force particles in 10-dimensional string theory, as well as in the newer 11-dimensional M-theory. The *Scientific American* article is deficient in details about the octonions; there is more in the online interview of Baez, and Baez’s *Bulletin* article explores the topic at an advanced level. Neither string theory nor M-theory offer any experimentally testable predictions, but the octonions are ready whenever string theory may be.

Shubin, Tatiana, David F. Hayes, and Gerald L. Alexanderson (eds.), *Expeditions in Mathematics*, MAA, 2011; xiv + 312 pp, $60.95 (MAA member: $48.95). ISBN 978-0-88385-571-3.

This book consists of 20 short articles on talks in the bimonthly Bay Area Mathematical Adventures series given since 2001 (earlier talks are in the MAA volume *Adventures for Students and Amateurs*, 2005). Many are by famous mathematicians and all are by gifted expositors of mathematics. Topics include mathematical “impossibilities,” the mathematics of sudoku, the Riemann hypothesis, soap bubbles, Hamiltonian circuits, and contributions of mathematics to drug therapies for HIV and leukemia.

Roy, Ranjan, *Sources in the Development of Mathematics: Series and Products from the Fifteenth to the Twenty-first Century*, Cambridge University Press, 2011; xv + 974 pp, $99. ISBN 978-0-521-11470-7.

I will mention this book, but I cannot review it. It is by a colleague in the office next to mine (and in its production I had a very small role). Here is what the book offers, excerpted from the back cover: “. . . many facets of the discovery and use of infinite series and products. . . context and motivation for these discoveries, including original notation and diagrams when practical. . . multiple derivations and. . . interesting exercises. . . .”

NEWS AND LETTERS

40th USA Mathematical Olympiad
2nd USA Junior Mathematical Olympiad

JACEK FABRYKOWSKI
Youngstown State University
Youngstown, OH 44555
jfabrykowski@ysu.edu

STEVEN R. DUNBAR
MAA American Mathematics Competitions
University of Nebraska-Lincoln
Lincoln, NE 68588-0658
sdunbar@maa.org

This year the Committee on the American Mathematics Competitions offered the USA Junior Mathematical Olympiad (USAJMO) for the second time, for students in 10th grade and below. Last year's experience shows that it provides a nicely balanced link between the computational character of the AIME problems and proof-oriented problems of USAMO. The USA Junior Mathematical Olympiad contained three problems on each of two days, with an allowed time of 4.5 hours each day—the same as the USAMO. Problems JMO1, JMO3 on Day 1, and problems JMO4, JMO5 were different from the USAMO problems, but JMO2 and JMO6 were the same as USAMO1 and USAMO4 respectively.

USAMO Problems

1. Let a, b, c be positive real numbers such that $a^2 + b^2 + c^2 + (a + b + c)^2 \le 4$. Prove that

$$\frac{ab+1}{(a+b)^2} + \frac{bc+1}{(b+c)^2} + \frac{ca+1}{(c+a)^2} \ge 3.$$

2. An integer is assigned to each vertex of a regular pentagon so that the sum of the five integers is 2011. A turn of a solitaire game consists of subtracting an integer m from each of the integers at two neighboring vertices and adding $2m$ to the opposite vertex, which is not adjacent to either of the first two vertices. (The amount m and the vertices chosen can vary from turn to turn.) The game is won at a certain vertex if, after some number of turns, that vertex has the number 2011 and the other four vertices have the number 0. Prove that for any choice of the initial integers, there is exactly one vertex at which the game can be won.

3. In the nonconvex but non-self-intersecting hexagon $ABCDEF$, no two opposing sides are parallel. The internal angles satisfy $\angle A = 3\angle D$, $\angle C = 3\angle F$, and $\angle E = 3\angle B$. Also, $AB = DE$, $BC = EF$, and $CD = FA$. Prove that the diagonals AD, BE, CF are concurrent.

Math. Mag. **84** (2011) 306–315. doi:10.4169/math.mag.84.4.306.

4. Consider the assertion that for each positive integer $n \geq 2$, the remainder upon dividing 2^{2^n} by $2^n - 1$ is a power of 4. Either prove the assertion or find (with proof) a counterexample.
5. Let P be a given point inside quadrilateral $ABCD$. Points Q_1 and Q_2 are located within $ABCD$ such that

$$\angle Q_1BC = \angle ABP, \qquad \angle Q_1CB = \angle DCP,$$
$$\angle Q_2AD = \angle BAP, \qquad \angle Q_2DA = \angle CDP.$$

   Prove that $\overline{Q_1Q_2} \parallel \overline{AB}$ if and only if $\overline{Q_1Q_2} \parallel \overline{CD}$.
6. Let A be a set with $|A| = 225$, meaning that A has 225 elements. Suppose further that there are eleven subsets $A_1, \ldots, A_{11}$ of A such that $|A_i| = 45$ for $1 \leq i \leq 11$ and $|A_i \cap A_j| = 9$ for $1 \leq i < j \leq 11$. Prove that $|A_1 \cup A_2 \cup \cdots \cup A_{11}| \geq 165$, and give an example for which equality holds.

USAMO Solutions

1. Set $2x = a + b$, $2y = b + c$, and $2z = c + a$; that is, $a = z + x - y$, $b = x + y - z$, and $c = y + z - x$. Hence

$$\frac{ab+1}{(a+b)^2} = \frac{(z+x-y)(x+y-z)+1}{4x^2}$$
$$= \frac{x^2-(y-z)^2+1}{4x^2} = \frac{x^2+2yz+1-y^2-z^2}{4x^2}.$$

   On the other hand, the given condition is equivalent to $2a^2 + 2b^2 + 2c^2 + 2ab + 2bc + 2ca \leq 4$ or $(a+b)^2 + (b+c)^2 + (c+a)^2 \leq 4$; that is, $x^2 + y^2 + z^2 \leq 1$ or $1 - y^2 - z^2 \geq x^2$. It follows that

$$\frac{ab+1}{(a+b)^2} = \frac{x^2+2yz+1-y^2-z^2}{4x^2} \geq \frac{x^2+2yz+x^2}{4x^2} = \frac{1}{2} + \frac{yz}{2x^2}.$$

   Likewise, we have

$$\frac{bc+1}{(b+c)^2} \geq \frac{1}{2} + \frac{zx}{2y^2} \quad \text{and} \quad \frac{ca+1}{(c+a)^2} \geq \frac{1}{2} + \frac{xy}{2z^2}.$$

   Adding the last three inequalities gives

$$\frac{ab+1}{(a+b)^2} + \frac{bc+1}{(b+c)^2} + \frac{ca+1}{(c+a)^2} \geq \frac{3}{2} + \frac{yz}{2x^2} + \frac{zx}{2y^2} + \frac{xy}{2z^2} \geq 3,$$

   by the AM–GM inequality. Equality holds if and only if $x = y = z$ or $a = b = c = \frac{1}{\sqrt{3}}$.

   This problem was suggested by Titu Andreescu. The solution is from Zuming Feng.
2. Let a_1, a_2, a_3, a_4, and a_5 represent the integers at vertices v_1 to v_5 (in order around the pentagon) at the start of the game. We will first show that the game can be won at only one of the vertices. Observe that the quantity $a_1 + 2a_2 + 3a_3 + 4a_4 \bmod 5$ is an invariant of the game. For instance, one move involves replacing a_1, a_3, and a_5 by $a_1 - m$, $a_3 + 2m$, and $a_5 - m$. Thus the quantity $a_1 + 2a_2 + 3a_3 + 4a_4$ becomes

$$(a_1 - m) + 2a_2 + 3(a_3 + 2m) + 4a_4 = a_1 + 2a_2 + 3a_3 + 4a_4 + 5m,$$

which is unchanged mod 5. The other moves may be checked similarly. Now suppose that the game may be won at vertex v_j. The value of the invariant at the winning position is $2011j$. If the initial value of the invariant is n, then we must have $2011j \equiv n \bmod 5$, or $j \equiv n \bmod 5$. Hence the game may only be won at vertex v_j, where j is the least positive residue of n mod 5.

By renumbering the vertices, we may assume without loss of generality that the winning vertex is v_5. We will show that the game can be won in four moves by adding a suitable amount $2m_j$ at vertex v_j (and subtracting m_j from the opposite vertices) on the jth turn for $j = 1, 2, 3, 4$. The net change at vertex v_1 after these four moves is $2m_1 - m_3 - m_4$, which must equal $-a_1$ if we are to finish with 0 at v_1. In this fashion we find that

$$\begin{aligned} 2m_1 - m_3 - m_4 &= -a_1 \\ 2m_2 - m_4 \qquad &= -a_2 \\ 2m_3 \qquad - m_1 &= -a_3 \\ 2m_4 - m_1 - m_2 &= -a_4 \\ -m_2 - m_3 &= -a_5 + 2011. \end{aligned}$$

The sum of the first four equations is the negative of the fifth equation, so the fifth equation is redundant. Multiplying the first four equations by $-1, 3, -3, 1$ and adding them yields $5m_2 - 5m_3 = a_1 - 3a_2 + 3a_3 - a_4$. But

$$a_1 - 3a_2 + 3a_3 - a_4 \equiv a_1 + 2a_2 + 3a_3 + 4a_4 \equiv n \equiv 5 \equiv 0 \bmod 5,$$

since we are assuming v_5 is the winning vertex. Therefore we may divide by 5 to obtain $m_2 - m_3 = \frac{1}{5}(a_1 - 3a_2 + 3a_3 - a_4)$. We also know that $m_2 + m_3 = a_1 + a_2 + a_3 + a_4$, and one easily confirms that the right-hand sides of these equations are integers with the same parity. Hence the system admits an integral solution for m_2 and m_3. The second and third equations then quickly give integer values for m_1 and m_4 as well, so it is indeed possible to win the game at vertex v_5.

This problem was suggested by Sam Vandervelde.

3. We first give a recipe for constructing hexagons as in the problem statement. Let ACE be a triangle, with all angles less than $2\pi/3$. Let D be the reflection of A across CE; let F be the reflection of C across EA; let B be the reflection of E across AC. Then, $\angle BAF = \angle BAC + \angle CAE + \angle EAF = 3\angle CAE = 3\angle CDE$, and similarly for the other angle equalities. Also, $AB = AE = DE$, and similarly for the other side equalities. Thus, the hexagon satisfies the equations in the problem statement. The diagonals AD, BE, CF are simply the altitudes of the triangle ACE, so they are concurrent at the orthocenter.

Now we show that the only possible hexagons meeting the conditions of the problem statement are the ones constructed in this manner. This will suffice to complete the solution.

Given the hexagon $ABCDEF$ as in the problem statement, let β, δ, ϕ be the measures of its angles B, D, F. Since $4(\angle B + \angle D + \angle F) = \angle A + \angle B + \angle C + \angle D + \angle E + \angle F = 4\pi$, we must have $\beta + \delta + \phi = \pi$. Also, the fact that opposite sides are not parallel implies that $\pi + 2\beta = \angle D + \angle E + \angle F \neq 2\pi$, so $\beta \neq \pi/2$; likewise $\delta, \phi \neq \pi/2$.

We can construct a hexagon $A_1B_1C_1D_1E_1F_1$ meeting the angle and side equality conditions, with angles $\angle B_1 = \beta$, $\angle D_1 = \delta$, $\angle F_1 = \phi$, by taking $A_1C_1E_1$ to be a triangle with angles β, δ, ϕ, and reflecting each vertex across the opposite side as above. We wish to show that $ABCDEF \sim A_1B_1C_1D_1E_1F_1$.

Treat the positions of A, B as fixed, and treat β, δ, ϕ as fixed; these are enough to uniquely determine the orientation of each edge of the hexagon, given the known angles. Let $x = AB = DE$, $y = BC = EF$, $z = CD = FA$. Our goal is to show that these lengths are uniquely determined (up to scale) by the given angles.

Let a, b, c, d, e, f be unit vectors in the directions of the edges from A to B, B to C, C to D, D to E, E to F, and F to A, respectively. Then the vector identity

$$x(a + d) + y(b + e) + z(c + f) = 0 \tag{1}$$

holds. Without loss of generality, assume the vertices of $ABCDEF$ are labeled in counterclockwise order. The respective orientations of vectors b, c, d, e, f, measured counterclockwise relative to a, are

$$b : \pi - \beta$$
$$c : -\beta - 3\phi$$
$$d : -2\phi$$
$$e : \pi + 2\delta - \beta$$
$$f : 2\delta - \phi - \beta$$

(These angles are given modulo 2π; we have made liberal use of the identity $\beta + \delta + \phi = \pi$.)

Now, whenever two unit vectors point in directions θ and ψ, which do not differ by π, then their sum is a nonzero vector pointing in direction $(\theta + \psi)/2$ or $(\theta + \psi)/2 + \pi$. It follows that vectors $a + d$, $b + e$, $c + f$ are all nonzero and point in the following directions (modulo π):

$$a + d : -\phi$$
$$b + e : \delta - \beta$$
$$c + f : \delta - 2\phi - \beta$$

None of the differences between these angles are multiples of π. (This follows from the fact that $\beta, \delta, \phi \neq \pi/2$.) Thus, $a + d$, $b + e$, $c + f$ are not collinear. Consequently, the equation (1) determines the coefficients x, y, z uniquely up to scale, as required.

It follows that $ABCDEF$ and $A_1B_1C_1D_1E_1F_1$ are similar to each other, as required, and this completes the proof.

This problem was suggested by Gabriel Carroll.

4. The assertion is false, and the smallest n for which it fails is $n = 25$. Given $n \geq 2$, let r be the remainder when 2^n is divided by n. Then $2^n = kn + r$ where k is a positive integer and $0 \leq r < n$. It follows that

$$2^{2^n} = 2^{kn+r} \equiv 2^r \mod 2^n - 1,$$

and $2^r < 2^n - 1$ so 2^r is the remainder when 2^{2^n} is divided by $2^n - 1$. If r is even then 2^r is power of 4. Hence to disprove the assertion, it is enough to find an n for which the corresponding r is odd.

If n is even then so is $r = 2^n - kn$.

If n is an odd prime then $2^n \equiv 2 \mod n$ by Fermat's Little Theorem; hence $r \equiv 2^n \equiv 2 \mod n$ and $r = 2$.

There remains the case in which n is odd and composite. In the first three instances $n = 9, 15, 21$ there is no contradiction to the assertion:

$$n = 9 : 2^6 \equiv 1 \mod 9 \Rightarrow 2^9 \equiv 2^6 \cdot 2^3 \equiv 8 \mod 9$$
$$n = 15 : 2^4 \equiv 1 \mod 15 \Rightarrow 2^{15} \equiv (2^4)^3 \cdot 2^3 \equiv 8 \mod 15$$
$$n = 21 : 2^6 \equiv 1 \mod 21 \Rightarrow 2^{21} \equiv (2^6)^3 \cdot 2^3 \equiv 8 \mod 21$$

However,

$$2^{10} = 1024 \equiv -1 \Rightarrow 2^{20} \equiv 1 \Rightarrow 2^{25} \equiv 2^5 \equiv 7 \mod 25,$$

so 7 is the remainder when 2^{25} is divided by 25 and 2^7 is the remainder when $2^{2^{25}}$ is divided by $2^{25} - 1$.

This problem was suggested by Sam Vandervelde.

5. We will prove that the lines $\overline{AB}$, $\overline{CD}$, and $\overline{Q_1Q_2}$ are either concurrent or all parallel. Let X and Y denote the reflections of P across the lines $\overline{AB}$ and $\overline{CD}$. We first claim that $XQ_1 = YQ_1$ and $XQ_2 = YQ_2$. Indeed, let Z be the reflection of Q_1 across BC. Then $XB = PB$, $BQ_1 = BZ$, and

$$\angle XBQ_1 = \angle XBA + \angle ABQ_1 = \angle ABC = \angle PBC + \angle CBZ = \angle PBZ,$$

whence $\triangle XBQ_1 \cong \triangle PBZ$ and thus $XQ_1 = PZ$. Similarly $YQ_1 = PZ$, and so $XQ_1 = YQ_1$. In exactly the same way, we see that $XQ_2 = YQ_2$, establishing the claim.

We conclude that the line $\overline{Q_1Q_2}$ is the perpendicular bisector of the segment $\overline{XY}$. If $\overline{AB} \parallel \overline{CD}$, then $\overline{XY} \perp \overline{AB}$ and it follows that $\overline{Q_1Q_2} \parallel \overline{AB}$, as desired. If the lines $\overline{AB}$ and $\overline{CD}$ are not parallel, then let R denote their intersection. Since $RX = RP = RY$, R lies on the perpendicular bisector of $\overline{XY}$ and thus R, Q_1, Q_2 are collinear, as desired.

This problem was suggested by Delong Meng and Zuming Feng.

6. Let S be the complement of $A_1 \cup A_2 \cup \cdots \cup A_{11}$ in A; we wish to prove that $|S| \le 60$. For $\ell \ge 0$, define

$$\theta(\ell) = \left(1 - \frac{\ell}{2}\right)\left(1 - \frac{\ell}{3}\right) = 1 - \frac{2}{3}\ell + \frac{1}{3}\binom{\ell}{2}.$$

Note that $\theta(0) = 1$ and $\theta(\ell) \ge 0$ for any integer $\ell > 0$. Therefore, since S is the intersection of the complements of the A_i,

$$|S| \le \sum_{n \in A} \theta(\ell(n)).$$

On the other hand,

$$\sum_{n\in A} \theta(\ell(n)) = \sum_{n \in A}\left(1 - \frac{2}{3}\ell(n) + \frac{1}{3}\binom{\ell(n)}{2}\right)$$
$$= |A| - \frac{2}{3}\sum_i |A_i| + \frac{1}{3}\sum_{i<j} |A_i \cap A_j|.$$

Consequently,

$$|S| \le 225 - \frac{2}{3} \cdot 11 \cdot 45 + \frac{1}{3} \cdot \binom{11}{2} \cdot 9 = 60,$$

and therefore $|A_1 \cup A_2 \cup \cdots \cup A_{11}| \ge 165$.

We construct an example to show that this lower bound is best possible. Let $p_1, p_2, \ldots, p_{11}$ be a set of 11 distinct primes, and let A' denote the set of all products of three of these primes. Furthermore, let $A'' = \{q_1, q_2, q_3, \ldots, q_{60}\}$ be a set of 60 distinct positive integers that are all prime to $p_1, \ldots, p_{11}$. Set $A = A' \cup A''$, and define

$$A_i = \{n \in A' : p_i \mid n\}.$$

Then $|A_i| = \binom{10}{2} = 45$, $|A_i \cap A_j| = \binom{9}{1} = 9$, and

$$|A_1 \cup A_2 \cup \cdots \cup A_{11}| = |A'| = \binom{11}{3} = 165.$$

Finally, $|A| = |A'| + |A''| = 165 + 60 = 225$.

This problem was suggested by Sid Graham.

USAJMO Problems

1. Find, with proof, all positive integers n for which $2^n + 12^n + 2011^n$ is a perfect square.
2. Same as USAMO 1.
3. For a point $P = (a, a^2)$ in the coordinate plane, let $\ell(P)$ denote the line passing through P with slope $2a$. Consider the set of triangles with vertices of the form $P_1 = (a_1, a_1^2)$, $P_2 = (a_2, a_2^2)$, $P_3 = (a_3, a_3^2)$, such that the intersections of the lines $\ell(P_1)$, $\ell(P_2)$, $\ell(P_3)$ form an equilateral triangle Δ. Find the locus of the center of Δ as $P_1P_2P_3$ ranges over all such triangles.
4. A *word* is defined as any finite string of letters. A word is a *palindrome* if it reads the same backwards as forwards. Let a sequence of words $W_0, W_1, W_2, \ldots$ be defined as follows: $W_0 = a$, $W_1 = b$, and for $n \geq 2$, W_n is the word formed by writing W_{n-2} followed by W_{n-1}. Prove that for any $n \geq 1$, the word formed by writing $W_1, W_2, \ldots, W_n$ in succession is a palindrome
5. Points A, B, C, D, E lie on circle ω and point P lies outside the circle. The given points are such that (i) lines PB and PD are tangent to ω, (ii) P, A, C are collinear, and (iii) $\overline{DE} \parallel \overline{AC}$. Prove that $\overline{BE}$ bisects $\overline{AC}$.
6. Same as USAMO 4.

Solutions

1. The answer is $n = 1$. Clearly, $n = 1$ is a solution because $2 + 12 + 2011 = 45^2$. Next we show that there is no other solution.

   Assume that $n \geq 2$. If n is odd, then $2^n + 12^n + 2011^n$ cannot be a perfect square because it is congruent to 3 modulo 4. If n is even, we can complete our solution in two ways.

   - $2^n + 12^n + 2011^n$ cannot be a perfect square because it is congruent to 2 modulo 3.
   - $2^n + 12^n + 2011^n$ cannot be a perfect square because it is between two consecutive perfect squares. Indeed, say $n = 2k$, then

$$\begin{aligned}(2011^k)^2 &< 2^{2k} + 12^{2k} + 2011^{2k} = 4^k + 144^k + 2011^{2k} \\ &< 1 + 2 \cdot 2011^k + 2011^{2k} = (2011^k + 1)^2.\end{aligned}$$

   This problem was suggested by Titu Andreescu.

3. For $1 \le i < j \le 3$, solving the system $y = 2x_i x - x_i^2 = 2x_j x - x_j^2$ yields the intersection $\left(\frac{x_i+x_j}{2}, x_i x_j\right)$ of lines ℓ_i and ℓ_j. Hence the center of the equilateral triangle is

$$O = (O_x, O_y) = \left(\frac{x_1 + x_2 + x_3}{3}, \frac{x_1x_2 + x_2x_3 + x_3x_1}{3}\right).$$

Let $0^\circ \le \alpha_i < 180^\circ$ be the standard angle formed by lines ℓ_i and the positive x-axis. Without loss of generality, we may assume that $\alpha_1 < \alpha_2 < \alpha_3$. By the given condition, we have $\alpha_2 - \alpha_1 = \alpha_3 - \alpha_2 = 60^\circ$. By the subtraction formulas, we have

$$\tan 60^\circ = \frac{\tan\alpha_2 - \tan\alpha_1}{1 + \tan\alpha_1 \tan\alpha_2} = \frac{\tan\alpha_3 - \tan\alpha_2}{1 + \tan\alpha_2 \tan\alpha_3}$$
$$\tan 120^\circ = \frac{\tan\alpha_3 - \tan\alpha_1}{1 + \tan\alpha_3 \tan\alpha_1}$$

or

$$\sqrt{3} = \frac{2x_2 - 2x_1}{1 + 4x_1x_2} = \frac{2x_3 - 2x_2}{1 + 4x_2x_3} \quad \text{and} \quad -\sqrt{3} = \frac{2x_3 - 2x_1}{1 + 4x_3x_1}.$$

Therefore,

$$1 + 4x_1x_2 = \frac{2(x_2 - x_1)}{\sqrt{3}}, \quad 1 + 4x_2x_3 = \frac{2(x_3 - x_2)}{\sqrt{3}},$$
$$1 + 4x_3x_1 = \frac{2(x_1 - x_3)}{\sqrt{3}}. \tag{2}$$

Adding these equations gives $3 + 4(x_1x_2 + x_2x_3 + x_3x_1) = 0$, implying that $O_y = -\frac{1}{4}$; that is, O always lies on the directrix ℓ of the parabola $y = x^2$.

Next we show that G can be any point on ℓ. Solving the first and third equations in (2) for x_2 and x_3 in terms of x_1 gives

$$x_2 = \frac{2x_1 + \sqrt{3}}{2 - 4\sqrt{3}x_1} \quad \text{and} \quad x_3 = \frac{2x_1 - \sqrt{3}}{2 + 4\sqrt{3}x_1},$$

implying that

$$x_1 + x_2 + x_3 = x_1 + \frac{(2x_1 + \sqrt{3})(2 + 4\sqrt{3}x_1) + (2x_1 - \sqrt{3})(2 - 4\sqrt{3}x_1)}{4 - 48x_1^2}$$
$$= x_1 + \frac{8x_1}{1 - 12x_1^2} = \frac{12x_1^3 - 9x_1}{12x_1^2 - 1}.$$

Because lines ℓ_1, ℓ_2, ℓ_3 are evenly spaced with 60° between each other, slopes $2x_1, 2x_2, 2x_3$ are symmetric with each other; that is,

$$x_1 + x_2 + x_3 = \frac{12x_i^3 - 9x_i}{12x_i^2 - 1} \quad \text{for } i = 1, 2, 3.$$

Therefore,

$$O_x = \frac{x_1 + x_2 + x_3}{3} = \frac{4x^3 - 3x}{12x^2 - 1},$$

where $-\infty < x < \infty$, because $x = x_i$ for some $i = 1, 2, 3$, and the combined ranges of slopes $2x_i$ are the interval $(-\infty, \infty)$. Because $4x^3 - 3x = O_x(12x^2 - 1)$ is a cubic equation, it has a real root in x for every real number O_x; that is, the range of O_x is the interval $(-\infty, \infty)$. We conclude that the locus of O is line $y = -\frac{1}{4}$.

This problem was suggested by Zuming Feng.

4. According to the statement of the problem we have

$$W_0 = a, \quad W_1 = b, \quad W_2 = ab, \quad W_3 = bab, \quad W_4 = abbab,$$

and so forth. Let $V_n = W_1 W_2 \cdots W_n$, where we place two or more words next to one another to denote the single word obtained by writing all their letters in succession. We find that

$$V_1 = b, \quad V_2 = bab, \quad V_3 = babbab, \quad V_4 = babbababbab.$$

We wish to show that V_n is a palindrome for all positive integers n. The above list shows this to be true for $1 \leq n \leq 4$; these cases will serve as the base cases for a proof by strong induction.

We use a bar over a word to indicate writing its letters in the reverse order. Thus $\overline{W_4} = babba$ and $\overline{V_3} = V_3$ since V_3 is a palindrome. Now assume that the words V_1 through V_n are all palindromes; we will show that V_{n+1} is also a palindrome. By the definition of V_{n+1} and W_{n+1} we have

$$V_{n+1} = V_n W_{n+1} = \overline{V_n}\, W_{n-1} W_n,$$

using the fact that $\overline{V_n} = V_n$ since V_n is a palindrome. But we know that $V_n = V_{n-2} W_{n-1} W_n$, so we may write

$$\overline{V_n}\, W_{n-1} W_n = \overline{W_n}\, \overline{W_{n-1}}\, \overline{V_{n-2}}\, W_{n-1} W_n.$$

The latter word is clearly a palindrome since V_{n-2} reads the same forward as backwards. Hence V_{n+1} is a palindrome, thus completing the proof.

This problem was suggested by Gabriel Carroll.

5. Let O be the center of circle ω and let M be the midpoint of $\overline{AC}$. It is clear that $\overline{BE}$ bisects $\overline{AC}$ if and only if E, M, B are collinear. Consequently, it suffices to show that

$$\angle MED = \angle BED. \tag{3}$$

The proof is divided into four parts.

(a) Triangle MED is isosceles with $\angle MED = \angle MDE$. (Note that $ACDE$ is an isosceles trapezoid and M is midpoint of the base $\overline{AC}$. The fact that triangle MED is isosceles then follows by the Pythagorean Theorem if nothing more elegant comes to mind.) This fact together with Alternate Interior Angles gives

$$\angle AME = \angle MED = \angle MDE = \angle PMD.$$

(b) *Claim.* The circle ω' with diameter $\overline{OP}$ contains points B, D, and M.

Proof. For each of the cases $X = B, D, M$, it is straightforward to verify that $\overline{OX}$ is perpendicular to $\overline{PX}$. For $X = B$ it is true that $\overline{OBP}$ is a right angle because $\overline{PB}$ is tangent to the circle at B. The same is true for $X = D$. For $X = M$, simply use the fact that if M is the midpoint of any given chord, then $\overline{OM}$ is perpendicular to the chord.

(c) Referring to the circle ω', the Inscribed Angle Theorem gives $\angle PBD = \angle PMD$.

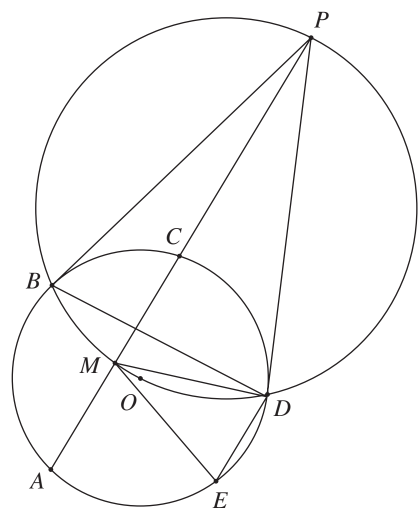
(d) Because $\overline{BP}$ is tangent to ω at B,

$$\angle BED = \frac{1}{2}\,\overset{\frown}{BD} = \angle PBD.$$

Results from step 1 yield

$$\angle BED = \angle PBD = \angle PMD = \angle MED,$$

establishing (3) and completing the proof.



This problem was suggested by Zuming Feng.

A total of 277 students participated in the USAMO on April 27 and 28, 2011. The top twelve students on the 2011 USAMO were (in alphabetical order):

Wenyu Cao	12	Phillips Academy	Andover	MA
Zijing Gao	9	Cary Academy	Cary	NC
Benjamin Gunby	11	Georgetown Day School	Washington	DC
Xiaoyu He	11	Acton-Boxborough Regional HS	Acton	MA
Ravi Jagadeesan	9	Phillips Exeter Academy	Exeter	NH
Yong Wook Kwon	11	Phillips Exeter Academy	Exeter	NH
Mitchell Lee	11	Thomas Jefferson High School	Alexandria	NH
Ray Li	10	Phillips Exeter Academy	Exeter	NJ
Evan O'Dorney	12	Berkeley Math Circle	Berkeley	CA
Mark Sellke	9	William Henry Harrison HS	West Lafayette	IN
David Yang	10	Phillips Exeter Academy	Exeter	NH
Shijie (Joy) Zheng	12	Phillips Exeter Academy	Exeter	NH

A total of 233 students participated in the USAJMO on April 27 and 28, 2011. The top fourteen students on the 2011 USAJMO were (in alphabetical order):

Evan Chen	9	Irvington HS	Fremont	CA
Eric Chen	10	Des Moines Central Academy	Des Moines	IA
Vahid Fazel Rezai	9	Red River HS	Grand Forks	ND
Owen Goff	9	Home School	Olympia	WA
Andrew He	8	John F Kennedy HS	Cupertino	CA
David Liang	9	Carmel HS	Carmel	IN
Aaron Lin	9	Mission San Jose HS	Fremont	CA
Jeffrey Ling	10	Palo Alto HS	Palo Alto	CA
Zhiyao Ma	10	Interlake HS	Bellevue	WA
Eric Neyman	8	Takoma Park MS	Silver Spring	MD
Sachin Pandey	8	Thomas S Wootton HS	Rockville	MD
Abram Sanderson	10	Wayzata HS	Plymouth	MN
Rachit Singh	9	Pullman HS	Pullman	WA
Kevin Zhou	10	High Technology HS	Lincroft	NJ

Evan O'Dorney and David Yang tied for the top score in this year's USA Mathematical Olympiad and both received the Samuel L. Greitzer/Murray S. Klamkin Award for Mathematical Excellence. Tom Ruff, Vice President of Akamai Technologies, presented O'Dorney and Yang with First Place Akamai Foundation Scholarship Awards. Xiaoyu He, Wenyu Cao, and Mitchell Lee received Second Place Akamai Foundation Scholarship Awards. The Wendy Ravech-Akamai Mathematics Scholar Award, presented for the first time, went to Shijie (Joy) Zheng. The twelve USAMO winners also received the Robert P. Balles Distinguished Mathematics Student Award to recognize and reward their high achievement in the world of mathematics competitions.

52nd International Mathematical Olympiad

ZUMING FENG
Phillips Exeter Academy
Exeter, NH 03833
zfeng@exeter.edu

YI SUN
Phillips Exeter Academy
Exeter, NH 03833
yisun@mit.edu

Problems (Day 1)

1. Given any set $A = \{a_1, a_2, a_3, a_4\}$ of four distinct positive integers, we denote the sum $a_1 + a_2 + a_3 + a_4$ by s_A. Let n_A denote the number of pairs (i, j) with $1 \leq i < j \leq 4$ for which $a_i + a_j$ divides s_A. Find all sets A of four distinct positive integers which achieve the largest possible value of n_A.
2. Let $\mathcal{S}$ be a finite set of at least two points in the plane. Assume that no three points of $\mathcal{S}$ are collinear. A *windmill* is a process that starts with a line ℓ going through a single point $P \in \mathcal{S}$. The line rotates clockwise about the *pivot* P until the first time that the line meets some other point belonging to $\mathcal{S}$. This point, Q, takes over as the new pivot, and the line now rotates clockwise about Q, until it next meets a point of $\mathcal{S}$. This process continues indefinitely, with the pivot always being a point from $\mathcal{S}$.

   Show that we can choose a point P in $\mathcal{S}$ and a line ℓ going through P such that the resulting windmill uses each point of $\mathcal{S}$ as a pivot infinitely many times.
3. Let f be a real-valued function defined on the set of real numbers that satisfies

$$f(x + y) \leq yf(x) + f(f(x))$$

   for all real numbers x and y. Prove that $f(x) = 0$ for all $x \leq 0$.

Problems (Day 2)

4. Let $n > 0$ be an integer. We are given a balance and n weights of weight 2^0, $2^1, \ldots, 2^{n-1}$. We are to place each of the n weights on the balance, one after another, in such a way that the right pan is never heavier than the left pan. At each step we choose one of the weights that has not yet been placed on the balance, and place it on either the left pan or the right pan, until all of the weights have been placed.

   Determine the number of ways in which this can be done.
5. Let f be a function from the set of integers to the set of positive integers. Suppose that, for any two integers m and n, the difference $f(m) - f(n)$ is divisible by $f(m - n)$. Prove that, for all integers m and n with $f(m) \leq f(n)$, the number $f(n)$ is divisible by $f(m)$.
6. Let ABC be an acute triangle with circumcircle Γ. Let ℓ be a tangent line to Γ, and let ℓ_a, ℓ_b and ℓ_c be the lines obtained by reflecting ℓ in the lines BC, CA and AB, respectively. Show that the circumcircle of the triangle determined by the lines ℓ_a, ℓ_b and ℓ_c is tangent to the circle Γ.

Math. Mag. **84** (2011) 316–319. doi:10.4169/math.mag.84.4.316.

Solutions

1. For any positive integer k, the sets $\{k, 5k, 7k, 11k\}$ and $\{k, 11k, 19k, 29k\}$ achieve the maximum value of $n_A = 4$.

   In general, let $A = \{a_1, a_2, a_3, a_4\}$ be labeled so that $0 < a_1 < a_2 < a_3 < a_4$. Then $a_2 + a_4$ and $a_3 + a_4$ are both strictly between $s_A/2$ and s_A, and so cannot divide s_A. Thus $n_A = 4$ is maximal. If the other pair-sums all divide s_A we must have

$$\begin{cases} a_2 + a_3 = (1/2)\ s_A & \text{(because } a_1 + a_4 \text{ also divides } s_A) \\ a_1 + a_3 = (1/n)\ s_A \\ a_1 + a_2 = (1/m)\ s_A \end{cases}$$

   where $2 < n < m$ (if $2 = n$ then $a_2 = a_1$ and if $n = m$ then $a_3 = a_2$). Subtracting the first equation from the sum of the last two gives $(1/n + 1/m - 1/2)\, s_a = 2a_1$, so that $1/n + 1/m > 1/2$. This equation can hold only if $n = 3$ and either $m = 4$ or $m = 5$. Now if $m = 4$ then $a_1 = (1/24)s_A$ and $A = \{k, 5k, 7k, 11k\}$; if $m = 5$, we have $a_1 = (1/60)s_A$ and $A = \{k, 11k, 19k, 29k\}$.

   This problem was proposed by Fernando Campos García of Mexico.

2. Call a point in $\mathcal{S}$ a *vertex*, and direct all lines so that they have right and left sides. Call a direction *ordinary* if no line with that direction passes through two vertices, and call a line ordinary if it has an ordinary direction. Let $n = |\mathcal{S}|$. Call a line a *balancing line* if it passes through exactly one vertex and has exactly $\lfloor (n-1)/2 \rfloor$ other vertices to its right.

   We first show that there exists an ordinary balancing line through any vertex P. Start with any ordinary directed line through P with, say, k points to its right. Rotating it 180° about P gives a line with $n - 1 - k$ points to its right. The number of points to the right of the ordinary lines in this process changes in increments of 1, so some ordinary line which occurs has exactly $\lfloor (n-1)/2 \rfloor$ points to its right. This is the desired ordinary balancing line.

   Now, choose any ordinary balancing line ℓ; we claim that the windmill starting from ℓ uses each vertex as a pivot infinitely often. In any windmill, the number of points to the right of the ordinary lines remains fixed, as at each pivot change the old and new pivots switch sides. Thus, because ℓ was balancing, each ordinary line in the windmill is balancing. Now, by definition, lines of each ordinary direction (which are hence balancing) occur infinitely often in this windmill. But there can be at most one balancing line in any direction, so the balancing lines we constructed through each vertex are the unique ones in their respective directions and must appear infinitely often, as needed.

   This problem was proposed by Geoff Smith of the United Kingdom.

3. Let $f(0) = a$ and $f(f(0)) = b$. Setting $x = 0$ in the given identity we obtain $f(y) \le ay + b$. Applying this to the last term of the given yields

$$f(x+y) \le yf(x) + af(x) + b. \tag{1}$$

   Substituting $x = z + a$, $y = -a$ gives $f(z) \le b$ for all $z \in \mathbb{R}$. Now applying this to the last term of the identity gives

$$f(x+y) \le yf(x) + b. \tag{2}$$

   Replacing x and y in (2) with $x + y$ and $-y$ gives $f(x) \le -yf(x+y) + b$. For $y < 0$, we can multiply (2) by $-y$ and add it to the last inequality, giving $f(x) \le -y^2 f(x) + b - yb$, or $f(x) \le b\left(\frac{1-y}{1+y2}\right)$ when $y < 0$. As $y \to -\infty$ the last expression gets arbitrarily close to zero, so we have $f(x) \le 0$ for all $x \in \mathbb{R}$.

Setting $x = 2a - 1$ and $y = 1 - a$ in (1) gives $f(2a - 1) \geq 0$. Since $f(x) \leq 0$ always, this means $f(2a - 1) = 0$. The given identity with $y = 0$ forces $f(x) \leq f(f(x))$ always, so $0 = f(2a - 1) \leq f(f(2a - 1)) = a$. This means that $a = f(0) = 0$ and thus $b = 0$ as well. Setting $y = -x$ in (2) and using these facts gives $0 \leq -xf(x)$ for all x. For $x < 0$, this implies that $f(x) \geq 0$; but we know $f(x) \leq 0$ always and $f(0) = 0$, so in fact $f(x) = 0$ whenever $x \leq 0$.

This problem was proposed by Igor Voronovich of Belarus. This solution is based on one by Oleg Golberg. (There are non-constant functions that satisfy the identity.)

4. The answer is $(2n - 1)!! = 1 \cdot 3 \cdot 5 \cdots (2n - 1)$.

   Call a sequence of moves *valid* if the right pan is never heavier than the left pan when making these moves. It suffices to give a $(2n + 1)$-to-1 mapping between valid sequences for weights $2^0, \ldots, 2^n$ and weights $2^0, \ldots, 2^{n-1}$.

   For a valid sequence of moves of weights $2^0, 2^1, \ldots, 2^n$, if we remove the move of putting weight 2^0 in this sequence and divide the remaining weights by 2, we obtain a valid sequence of moves of weights $2^0, \ldots, 2^{n-1}$. On the other hand, for a valid sequence S of weights $2^0, \ldots, 2^{n-1}$, doubling each weight gives a valid sequence S' of weights $2^1, \ldots, 2^n$. Note that the difference in weight between the left and right pans is always at least 2 after the first move in S'. Therefore, modifying S' by adding weight 2^0 to the left pan on the first move or to either pan on any move after the first yields $2n + 1$ valid sequences of weights $2^0, \ldots, 2^n$. These two constructions give the desired mapping.

   This problem was proposed by Morteza Saghafian of Iran.

5. Setting $n = 0$ in the given gives $f(m) \mid f(m) - f(0)$, hence $f(m) \mid f(0)$ for all m, while taking $m = 0$ yields $f(-n) \mid f(0) - f(n)$ for all n. Together, these show that $f(-n) \mid f(n)$ for all n, implying $f(n) = f(-n)$. It therefore suffices to show that for all $m, n > 0$ either $f(m) \mid f(n)$ or $f(n) \mid f(m)$.

   Assume the contrary and pick $m > n > 0$ violating the desired with $m + n$ minimal. Since $m - n > 0$ and $(m - n) + n = m < m + n$, the minimality of $m + n$ implies that either $f(n) \mid f(m - n)$ or $f(m - n) \mid f(n)$. If $f(n) \mid f(m - n)$, then $f(n) \mid f(m) - f(m - n)$ implies $f(n) \mid f(m)$, a contradiction. Therefore, $f(n) \nmid f(m - n)$, hence $f(m - n) \mid f(n)$ and $f(m - n) < f(n)$. Note that $f(m) \mid f(n) - f(n - m) = f(n) - f(m - n)$. Since $f(n) - f(m - n) > 0$, this means $f(m) < f(n)$. Now, by the given, we have $f(n) \mid f(m) - f(m - n)$, where $|f(m) - f(m - n)| < f(n)$ because $f(m), f(m - n) < f(n)$. Hence, it must be that $f(m) = f(m - n)$, implying $f(m) \mid f(n)$, a contradiction.

   This problem was proposed by Mahyar Sefidgaran of Iran. This solution is by Oleg Golberg.

6. Let A_1, B_1, C_1 be the intersections of ℓ_b and ℓ_c, ℓ_c and ℓ_a, and ℓ_a and ℓ_b. Let ℓ be tangent to Γ at T. Define points A_2, B_2, C_2 (distinct from A, B, C) on Γ so that $\widehat{TA} = \widehat{A_2A}$, $\widehat{TB} = \widehat{B_2B}$, and $\widehat{TC} = \widehat{C_2C}$. Let lines AB, BC, CA meet ℓ at C_3, A_3, B_3, respectively. Without loss of generality, we suppose that ℓ is such that B lies inside triangle $B_1A_3C_3$; other configurations are analogous. Let B_1B_2 intersect Γ again at H. We claim there is a homothety $\mathbf{H}$ centered at H sending $A_2B_2C_2$ to $A_1B_1C_1$ and Γ to the circumcircle Γ_1 of triangle $A_1B_1C_1$. Since H lies on Γ, such an $\mathbf{H}$ would show that Γ and Γ_1 are tangent.

   We first show that corresponding sides of $A_1B_1C_1$ and $A_2B_2C_2$ are parallel; by symmetry, it suffices to show that $B_2C_2 \parallel B_1C_1$. Let S be the intersection of lines B_2C_2 and ℓ. Since $\widehat{B_2T} = 2\widehat{BT}$ and $\widehat{TC_2} = 2\widehat{TC}$, we have $\angle B_2ST = \angle B_2C_2T - \angle STC_2 = 2(\angle BC_2T - \angle CTC_2)$. Because BC_2CT is cyclic, we have $2(\angle BC_2T - \angle CTC_2) = 2(\angle BCT - \angle CTC_2) = 2\angle BA_3T$. Because B_1A_3 and ℓ are reflections of each other across line BC, we have $2\angle BA_3T = \angle B_1A_3T$. Combining these equalities gives $\angle B_2ST = \angle B_1A_3T$, hence $B_2C_2 \parallel B_1C_1$.

It remains to show that A_1A_2 and C_1C_2 pass through H; by symmetry, it suffices to do so for C_1C_2. We claim first that the intersection I of B_1B and C_1C lies on Γ. Indeed, by definition A_1B_1, AB, ℓ concur at C_3, B_1C_1, BC, ℓ concur at A_3, and C_1A_1, CA, ℓ concur at B_3. By reflection properties, line AB (through C_3) bisects $\angle A_3C_3B_1$, and line BC (through A_3) bisects $\angle B_1A_3C_3$, so B is the incenter of triangle $B_1C_3A_3$ in our configuration. We see similarly that C is the excenter of triangle $C_1A_3B_3$. Computing, we see $\angle ABI = 180^\circ - \angle B_3BC_3 = 180^\circ - \left(90^\circ + \frac{\angle B_1A_3C_3}{2}\right) = 90^\circ - \frac{\angle B_1A_3C_3}{2}$ and $\angle ACI = \angle CB_3A_1 - \angle CC_1B_3 = \frac{\angle A_3B_3A_1}{2} - \frac{\angle A_3C_1B_3}{2} = \frac{\angle B_3A_3C_1}{2} = \frac{180^\circ - \angle B_1A_3C_3}{2} = 90^\circ - \frac{\angle B_1A_3C_3}{2}$. Hence, $\angle ACI = \angle ABI$, and $ACBI$ is cyclic, so I lies on Γ.

By Pascal's theorem on the (self-intersecting) cyclic hexagon B_2HC_2BIC, the intersection B_1 of B_2H and BI, the intersection X of C_2B and CB_2, and the intersection of HC_2 and IC all lie on B_1X. Now, because $\widehat{B_2B} = \widehat{BT}$ and $\widehat{C_2C} = \widehat{CT}$, CB_2 and BC_2 are the reflections of CT and BT across BC. Thus, their intersection X is the reflection of T across BC and lies on the reflection B_1C_1 of TS across BC. This means that B_1X is the same line as B_1C_1. Therefore, HC_2 passes through the intersection of IC and B_1C_1, which is C_1 because I lies on CC_1. Thus, C_1C_2 passes through H, as needed.

This problem was proposed by the Olympiad problem committee of Japan.

Results.

The IMO was held in Amsterdam, The Netherlands, on July 18–19, 2011. There were 564 competitors from 101 countries and regions. On each day contestants were given four and a half hours for three problems.

On this challenging exam, a perfect score was achieved by only one student, Lisa Sauermann (Germany). With this result, she becomes the most successful IMO participant of all time, having won 4 gold medals and 1 silver medal in her 5 participations. Each member of the USA team won a gold medal, ranking the USA 2nd among all 101 participating countries, behind China. This impressive performance is only the second time the entire USA team has won gold medals. The students' individual results were as follows.

- Wenyu Cao, who finished 12th grade at Phillips Academy in Andover, MA, won a gold medal.
- Ben Gunby, who finished 11th grade at Georgetown Day School in Washington, DC, won a gold medal.
- Xiaoyu He, who finished 11th grade at Acton-Boxborough Regional High School in Acton, MA, won a gold medal.
- Mitchell Lee, who finished 11th grade at Thomas Jefferson High School for Science and Technology in Alexandria, VA, won a gold medal.
- Evan O'Dorney from Danville, CA, who finished 12th grade (homeschooled through Venture School), won a gold medal.
- David Yang, who finished 10th grade at Phillips Exeter Academy in Exeter, NH, won a gold medal.

2011 Carl B. Allendoerfer Awards

The Carl B. Allendoerfer Awards, established in 1976, are made to authors of expository articles published in *Mathematics Magazine*. The Awards are named for Carl B. Allendoerfer, a distinguished mathematician at the University of Washington and President of the Mathematical Association of America, 1959–60.

Curtis D. Bennett, Blake Mellor, and Patrick D. Shanahan, "Drawing a Triangle on the Thurston Model of Hyperbolic Space," *Mathematics Magazine*, **83:2** (2010), p. 83–99. http://dx.doi.org/10.4169/002557010X482853

This skillfully written article carefully compares the classical Poincaré disk model of the hyperbolic plane with a physical paper model due to Thurston. The authors present a clear discussion of the differences between the models, including a comparison of lines in the Thurston model with true hyperbolic lines. They include new insights into a classical theorem from differential geometry.

The Thurston model is an approximation of the hyperbolic plane that is constructed by gluing together Euclidean "model triangles" in a non-standard way. Within this model there is a special class of geodesics known as "Thurston lines". The article begins with the question of what can be said about the angle sum of a "big" triangle bounded by Thurston lines. In their first proposition, the authors provide a formula for this angle sum. A surprising consequence of the formula is the conclusion that any such triangle can contain at most two model vertices within its interior. The easy-to-follow development of this result leads to an elegant discussion of the differences between the Thurston model and the Poincaré disk model. With the machinery developed for the discussion of triangles, the authors provide a hyperbolic version of Pick's Theorem, and explain why we should not be surprised by the result. The concluding sections deftly take the reader deeper by generalizing Thurston's model and presenting a version of the Gauss-Bonnet Formula without the usual heavy machinery of differential geometry.

The authors enhance their presentation with numerous colorful illustrations and encourage the reader to construct their own physical Thurston models with which to experiment as they read the article. The result is an engaging exposition that sheds new light on a fascinating classical topic and leads the reader to a concrete appreciation of the hyperbolic plane.

Joint Response from Curtis Bennett, Blake Mellor, and Patrick Shanahan. We are deeply honored to receive the Allendoerfer award from the Mathematical Association of America. This paper started as a conversation about a surprisingly difficult homework problem in a class for future elementary teachers and then took on a life of its own; we hope the results can be brought back to inspire future classroom discussions. We would like to thank our colleagues at Loyola Marymount University for their help and encouragement and our families: Jon and Sam Bennett; Nancy Balter and Eric and Clara Mellor; and Dana, Kasey and Cody Shanahan. We would finally like to thank the MAA and its membership for providing a rich community committed to excellence in mathematics, the teaching of mathematics, and mathematical exposition.

Math. Mag. **84** (2011) 320–322. doi:10.4169/math.mag.84.4.320.

Biographical Notes:

Curtis Bennett earned a B.S. in mathematics from Colorado State University in 1985 and received his Ph.D. in mathematics from the University of Chicago in 1990. Since graduating, he has taught at Michigan State University, Ohio State University, Bowling Green State University, and Loyola Marymount University. In 1993, Dr. Bennett was a co-founder of the Young Mathematicians Network and served on the editorial board of the YMN for three years. He works in a variety of fields, including the study of geometries associated to groups of Lie type, combinatorics, and the scholarship of teaching and learning. He was a 2000–2001 and 2003–2004 Carnegie Scholar with the Carnegie Foundation for the Advancement of Teaching, and he won the Haimo Award for Excellence in Teaching in 2010. In his spare time, he enjoys playing golf (badly), bicycling, and hiking.

Blake Mellor earned a B.A. in mathematics from Harvard University in 1993 and received his Ph.D. in mathematics from UC Berkeley in 1999, under the direction of Robion Kirby. After three years at Florida Atlantic University, he moved to Loyola Marymount University in Los Angeles, where he is currently an associate professor of mathematics. His research interests are in knot theory and spatial graphs, with occasional forays exploring connections between mathematics and the arts. He enjoys martial arts, ballroom dancing, science fiction, and playing with his children, Eric and Clara.

Patrick D. Shanahan earned a B.A. in mathematics from California State University, Long Beach in 1990. He attended graduate school at UC Santa Barbara where he completed an M.A. in 1992 and a Ph.D. in 1996, supervised by Daryl Cooper. A highlight in his graduate career was participating in the MSRI summer graduate student program in hyperbolic geometry led by David Epstein, Jane Gilman, and Bill Thurston. He joined the faculty at Loyola Marymount University in 1996 where he is currently a professor of mathematics. Professor Shanahan's main area of research is in geometric topology with an emphasis on knot theory. He has also co-authored the textbook *A First Course in Complex Analysis with Applications*, currently in its second edition. In his spare time you can find him at the beach surfing with his children, Kasey and Cody.

Gene Abrams and Jessica K. Sklar, "The Graph Menagerie: Abstract Algebra and the Mad Veterinarian," *Mathematics Magazine*, **83:3** (2010), p. 168–179. http://dx.doi.org/10.4169/002557010X494814

This accessible and well-written article begins with a fanciful concept from recreational mathematics: a machine that can transmogrify a single animal of a given species into a finite nonempty collection of animals from any number of species. Given this premise, a natural question arises: if a Mad Veterinarian has a finite slate of such machines, then which animal menageries are equivalent? To answer this question, the authors associate to the slate of machines a directed "Mad Vet" graph. They then show that the corresponding collection of equivalence classes of animal menageries forms a semigroup and use the structure of the Mad Vet graph to determine when this collection is actually a group. In addition, the authors show that the Mad Vet groups can be identified explicitly using the Smith normal form of a matrix closely related to the incidence matrix of the Mad Vet graph. Although these results nicely wrap up the core questions about Mat Vet scenarios, the authors go further by showing how generalizations of these ideas lead to current work about Leavitt path algebras and C^*-algebras.

This interesting article provides a wonderful example of what the authors call "cross-disciplinary pollination", using known results from one field to answer questions in another. Starting with a simply stated puzzle, the reader is led on an elegant tour of Mad Vet scenarios. Sights along the way include graph theory, equivalence relations, semigroups and the interplay between graph theory and group theory. The tour concludes with a glimpse of ongoing research in sophisticated algebraic and analytic mathematics.

Response from Gene Abrams and Jessica Sklar. We enjoyed collaborating on this article. Perhaps what pleases us the most is that our work with it touches on all three vertices of the faculty trinity. We were originally introduced to Mad Vet Puzzles while wearing our service hats at a workshop on how to create Math Teacher's Circles; with research hats on, we saw how these puzzles lead directly to cutting-edge questions in noncommutative ring theory and functional analysis. We have incorporated our work into our teaching, creating Mad Vet instructional materials which provide our students with new perspectives on equivalence relations and finitely generated abelian groups.

Many people played critical roles in bringing this article to fruition. We thank Enrique Pardo for contributing a 'first principles' proof of the result which provides the article's foundation. We are extremely grateful to *Mathematics Magazine* editors Frank Farris and Walter Stromquist, as well as to the anonymous referees: their valuable input allowed us to make lemonade out of the lemon that was the original version of the article. Finally, we give special thanks to Ken Ross for his sage advice and constant encouragement throughout the process.

We are thrilled and honored to receive the Allendoerfer Award.

Biographical Notes:

Gene Abrams is Professor of Mathematics at the University of Colorado Colorado Springs. He earned his Ph.D. in mathematics at the University of Oregon in 1981. He has been actively involved in mathematics-oriented community outreach K-12 educational activities. Gene has published research articles and lectured extensively (both in the U.S. and Europe) on topics in associative rings and their modules, focusing since 2005 on *Leavitt path algebras*, and has been honored with various teaching awards, including: 2002 Teacher of the Year (Mathematical Association of America Rocky Mountain Section), President's Teaching Scholar (University of Colorado system, a lifetime designation made in 1996), and 1988 Outstanding Teaching Award (UCCS campus).

When he's not riding his bicycle, Gene succumbs to his passions for baseball, skiing, and the *New York Times* Sunday Crossword. He has been married to Mickey since 1983; they have two children, Ben and Ellen.

Jessica K. Sklar is Associate Professor of Mathematics at Pacific Lutheran University in Tacoma, Washington. She received her Ph.D. in mathematics at the University of Oregon in 2001. She is enamored of recreational mathematics and previously published an article in *Mathematics Magazine* (79:5, 2006) on the use of linear algebra in solving computer game puzzles. She is passionate about teaching and about sharing the beauty of mathematics with lay readers; she won Pacific Lutheran University's Faculty Excellence in Teaching Award in 2005–2006. Jessica is currently co-editing a collection of essays on mathematics in popular culture with Elizabeth Sklar. She lives in Seattle with her three charming and impish cats.

MAA

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

CONTENTS

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS