LETTER FROM THE EDITOR

If you only know the Marquis de L'Hopital from the rule that bears his name, then have a look at our opening article for 2022. L'Hopital was the author of what is generally considered the first calculus textbook, and Fred Kuczmarski presents a convincing case that this work holds treasures for modern readers. In particular, he explains L'Hopital's exceedingly clever geometrical approach to the problem of constructing tangents to curves. Nowadays we tend to take calculus for granted, but the pioneers of the subject still have much to teach us.

The calculus theme continues with our next article, from Hans Musgrave and Ryan Zerr. They note that while Riemann based his approach to integration on filling out areas under curves with rectangles, it is also possible to use triangles as the approximating shape. A benefit of this approach, as explained so lucidly in their article, is that it leads to some interesting results on infinite series.

If you prefer discrete mathematics, then have a look at the article by Therese Aglialoro and Robert Hochberg. They take their inspiration from the venerable Rubik's cube. Specifically, they consider a dodecahedral variant of the cube known as the Megaminx. They consider the combinatorial problem of constructing various snake-like patterns on this puzzle. The investigation leads them to the problem of enumerating Hamilton circuits on the dual graph of the dodecahedron. I especially enjoyed reading their article since it is a reminder that interesting mathematics does not necessarily require dense thickets of notation.

Tien Chih and Demitri Plessas apply insights gleaned from Google's famous PageRank algorithm to the problem of ranking heavyweight boxers. Along the way they discuss topics from linear algebra, probability, and statistics. Brett Hemenway and David Hemenway make an interesting observation about measures of diversity in a population—that most individuals will find themselves among people who are "more like them" than the aggregate statistics indicate. Mathematically, this is reminiscent of the famous "friendship paradox" from sociology, which is the observation that most people have fewer friends than their friends have.

We round out our collection of articles with two shorter pieces. Şahin Koçak, Yunus Özdemir, and Gökçe Özkaya offer some insights into the p-adic numbers, as well as into subtle questions involving the Gromov-Hausdorff measure on compact metric spaces. And Quang Hung Tran brings us home with a new proof of Pitot's theorem from Euclidean geometry. That is the one that says a convex quadrilateral has an inscribed circle precisely when the lengths of its two pairs of opposite sides have the same sum. This result has been known since the eighteenth century, but Tran's proof is impressive for its novelty and clarity.

We also have proofs without words, problems, reviews, and a report on the 50th USA Mathematical Olympiad. Truly we are starting 2022 in style!

Jason Rosenhouse, Editor

Math. Mag. **95** (2022) 3. doi:10.1080/0025570X.2022.2001229

ARTICLES

L'Hôpital's *Analyse*: Seeing the Infinitely Small

FRED KUCZMARSKI
Shoreline Community College
Shoreline, WA 98133
fkuczmar@shoreline.edu

In telling the story of L'Hôpital's rule, it is easy to dismiss L'Hôpital as a mere patron. This is unfortunate. For starters, the Marquis de L'Hôpital did not really purchase the rule that bears his name. Rather, he paid Johannes Bernoulli to tutor him in the newly discovered calculus. But more to the point, in turning his notes from these lessons into the first text on differential calculus, L'Hôpital crafted a work that bears the stamp of his genius for both exposition and generalization.

L'Hôpital chose an apt title for his text. His *Analyse des Infiniments Petits* (*Analysis of the Infinitely Small*), first published anonymously in 1696, focuses not on the derivative, but on *differentials*. Some 156 figures capture the brilliance of L'Hôpital's arguments on a differential level. While this type of reasoning has all but disappeared from our calculus curriculum, my hope is that this article inspires you to pick up a copy of the recent English translation of the *Analyse* [**1**] and share the spirit of L'Hôpital's approach with your students.

Our focus here is on Chapter 2, *Use of the Differential Calculus for Finding the Tangents of All Kinds of Curved Lines*. Today we might consider the problem of finding tangents solved once our students have learned to differentiate. But the mathematicians of L'Hôpital's era were focused on geometry and "finding the tangent to a curved line" meant describing a way to construct the tangent; not necessarily to construct with straightedge and compass, but to describe a clear method of drawing the tangent.

We focus in particular on Proposition X, where L'Hôpital turns his attention to constructing *normal* lines to a broad class of curves that he calls generalized conics. He formulates Proposition X as a problem:

Problem. *Let AMB be a curved line such that if we draw the straight lines MF, MG, MH, etc., from any one of its points M to the foci F, G, H, etc., their relationship is expressed by any equation. We wish to draw MP, which is perpendicular to the tangent at the given point M*

This proposition in particular shows L'Hôpital's genius for generalization. We explore a few of its many applications and show its relation to the gradient vector, as well as to a chain rule in multivariable calculus.

Normals to a Cassini oval

To pique your curiosity, consider the Cassini oval of Figure 1(a). The oval is the set of points whose distances to the foci F and G have a constant product. The problem

Math. Mag. **95** (2022) 4–13. doi:10.1080/0025570X.2022.2001261

is to construct the normal to the oval at M. But to emphasize the geometric nature of this problem and give ourselves the mindset of a 17th century mathematician, let us remove both the coordinate system and the oval (see Figure 1(b)), and ask the same question. How would you construct the normal at M?

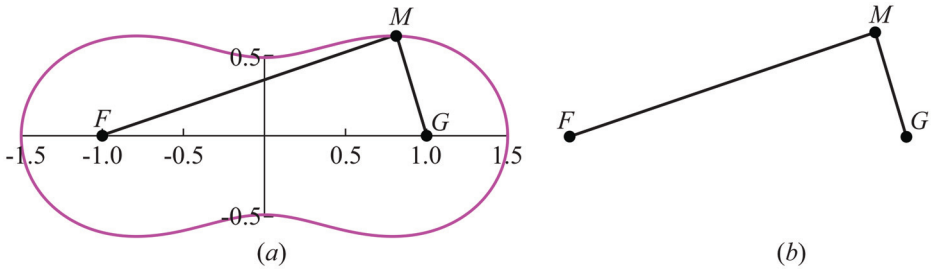


Figure 1 Construct the normal to the Cassini oval at M.

To start, let u and v be the respective distances from a point M of the oval to the foci F and G. Then the Cassini oval has equation $uv = k^2$ for some constant k, and the differential changes du and dv along the oval satisfy the relation

$$v\,du + u\,dv = 0. \tag{1}$$

To represent the differentials, L'Hôpital draws an "infinitely small" arc Mm on the curve and then constructs the "little circular arcs" MR and MS centered respectively at F and G (Figure 2). Then $Rm = du$ and $Sm = dv$. In the figure, $du > 0$ and $dv < 0$. These signs play an important role later on.

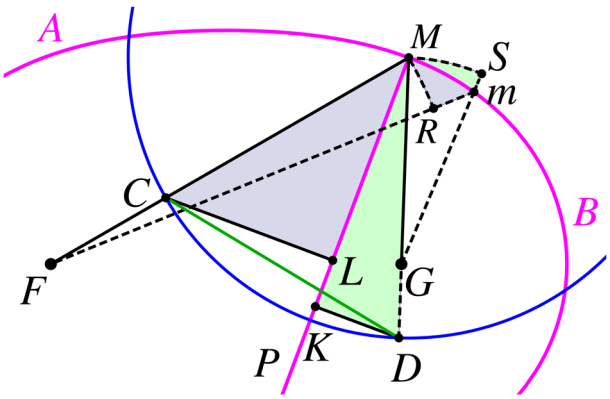


Figure 2 Constructing the normal (MP) at M to the curve AMB with foci F and G.

L'Hôpital next constructs *visible* copies of the differentials, taking care to scale these copies by the same factor. To establish the scaling factor, L'Hôpital constructs a circle of arbitrary radius centered at M, intersecting rays $\overrightarrow{MF}$ and $\overrightarrow{MG}$ at C and D. He then drops the perpendicular CL to the imagined, but yet to be constructed normal line MP. Subtracting the common angle $\angle LMR$ from the right angles $\angle LMm$ and $\angle RMC$ shows that $\angle RMm = \angle LMC$, and hence that $\triangle MRm \sim \triangle MLC$. We like to think of $\triangle MLC$ as a visible copy of the differential triangle $\triangle MRm$.

By dropping the perpendicular DK to MP, L'Hôpital also constructs a visible copy $\triangle MKD$ of the differential triangle $\triangle MSm$. Then as L'Hôpital explains:

> . . . because the hypotenuse Mm is common to the little triangles MRm and MSm and the hypotenuses MC and MD of the triangles MLC and MKD are equal to one another, it follows that the perpendiculars CL and DK have the same ratios among themselves as the differentials Rm and Sm.

So $CL : DK = Rm : Sm$ and "by substituting for Rm and Sm their proportional values CL and DK", L'Hôpital can rewrite the differential relation (1) as

$$v\,CL + u\,DK = 0. \tag{2}$$

This relationship is the key to his construction of the normal. Before continuing, we should remember that since the differentials in equation (1) are signed, so are the segments CL and DK. L'Hôpital is careful to point out, in what we call the *separation lemma*, that differentials corresponding to lengths from foci on opposite sides of the normal MP have opposite signs. In L'Hôpital's words:

> The lines that emanate from the foci situated on the same side of the perpendicular MP increase while the others decrease, or vice versa. For example, in Figure 2, FM increases by its differential Rm, while GM decreases by its differential Sm.*

Since $u, v > 0$ in equation (1), the differentials have opposite signs and the foci lie on opposite sides of the normal. This condition together with equation (2) defines the desired normal MP. In Figure 2, for example, $FM : GM = u : v = 2 : 1$, so the normal is the line that separates the foci and is twice as far from C as from D. You might pause here and consider how to construct this line.

To construct MP, L'Hôpital imagines the points C and D to be loaded with respective weights v and u, and interprets the expression $v\,CL + u\,DK$ as a sum of moments about the line MP. He then appeals to the principle of mechanics that

> . . . any straight line which passes through the center of gravity of several weights separates them so that the weights on one side multiplied by their distances from the line are precisely equal to the weights on the other side, each one also multiplied by its distance to the same straight line.

So, now with $u : v = 3 : 1$ as in Figure 3, L'Hôpital constructs the normal through the center of gravity X of weights w_C and $w_D = 3w_C$ at C and D, respectively. We need not rely on the principle of mechanics to see why this construction works. For since X is the point on segment CD with $CX : XD = 3 : 1$, it follows from similar triangles $\triangle CLX$ and $\triangle DKX$ that $CL : DK = 3 : 1$.

At times L'Hôpital uses Archimedes' law of the lever to modify his construction of the normal. We can apply this idea here to construct a normal to the Cassini oval without knowing the ratio $u : v$. To start, draw circles through F and G centered at M, and let C' and D' be the respective intersections of these circles with $\overrightarrow{MF}$ and $\overrightarrow{MG}$ (Figure 4). We claim the normal passes through the midpoint Z of segment $C'D'$. To see why, let's suppose as before that $u : v = 3 : 1$. Then by Proposition X, MP passes through the center of gravity Y of weights $w_{C'}$ and $w_G = 3w_{C'}$. Replacing the weight $w_G = 3w_{C'}$ at G with the weight $w_{C'}$ at D', three times as far from MP as G, leaves the moment about MP unchanged. Hence, the normal passes through the center of gravity Z of equal weights at C' and D'.

*Author's note: The differential du is positive when m and F lie on opposite sides of MP as in Figure 2.

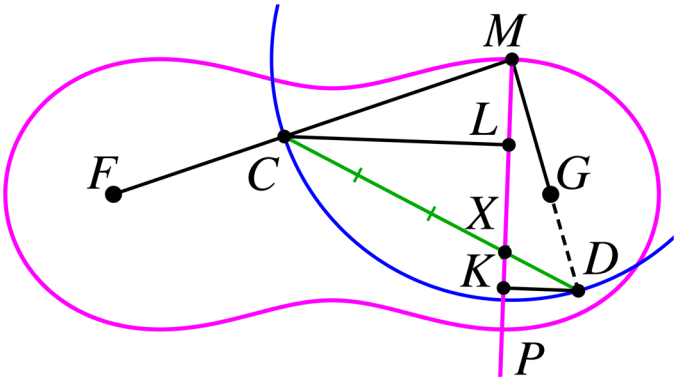


Figure 3 Constructing the normal through the center of gravity X.

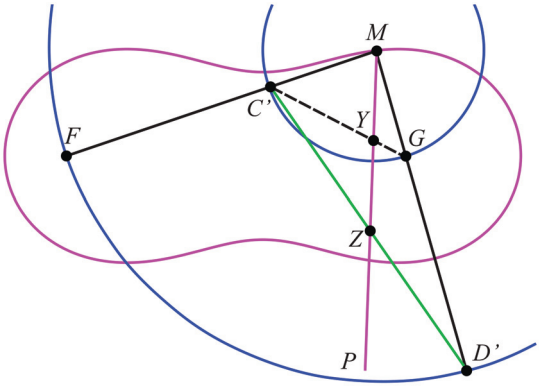


Figure 4 A modified construction of the normal using Archimedes' law of the lever.

We can express the same argument using vectors in place of weights. Writing

$$\overrightarrow{MY} = \frac{1}{4}\left(\overrightarrow{MC'} + 3\overrightarrow{MG}\right)$$

and

$$2\overrightarrow{MZ} = \overrightarrow{MC'} + \overrightarrow{MD'} = \overrightarrow{MC'} + 3\overrightarrow{MG} = 4\overrightarrow{MY},$$

shows that $\overrightarrow{MZ}$ is parallel to $\overrightarrow{MY}$ and hence normal to the oval.

Negative weights

The next example illustrates how L'Hôpital treats negative weights.

Circles of Apollonius. *The curve AMB in Figure 5 is defined as the set of points through M whose distances to the foci F and G have a constant ratio k. Draw the normal to AMB at M.*

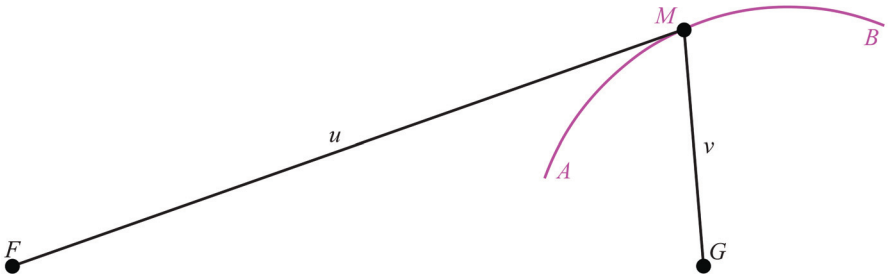


Figure 5 Construct the normal to the curve AMB with equation $u/v = k$ at M.

SOLUTION: With u and v representing the respective distances to F and G, an equation of the curve is $u/v = k$, implying that

$$v\,du - u\,dv = 0.$$

Instead of drawing an arbitrary circle centered at M, we draw the circle through G. Because the differential dv is loaded with a negative weight, L'Hôpital would construct the reflection G' of G in M and draw the normal through the center of gravity X of *positive* weights $w_{G'} : w_C = u : v = k$. We illustrate this construction in Figure 6 where $k = 3$. The differentials have the *same* sign; so the foci lie on the same side of MP and $CL : GK = 3 : 1$.

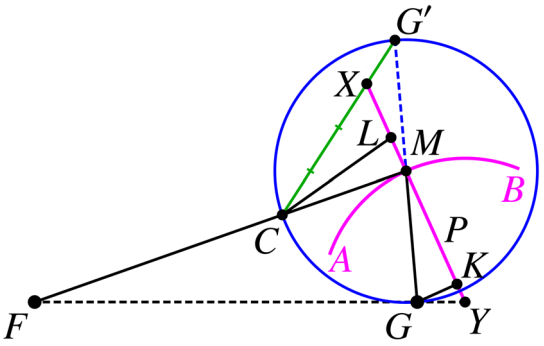


Figure 6 Constructing the normal at M for the curve $u/v = 3$.

When $k \neq 1$ the curve $u/v = k$ is a circle, and the law of the lever helps explain why. First, we allow for negative weights and construct MP through the center of gravity of weights $w_G = k^2$ and $w_C = -k$. Replacing the weight $-k$ at C with the weight -1 at F, k times as far from MP as C, leaves the moment about MP unchanged. Thus, MP passes through the center of gravity Y of weights $w_G = k^2$ and $w_F = -1$. But since

$$\overrightarrow{MY} = \left(k^2\overrightarrow{MG} - \overrightarrow{MF}\right)/(k^2 - 1),$$

the normals all pass through the *fixed* point Y on $\overleftrightarrow{FG}$, where $\overrightarrow{FY} = k^2\overrightarrow{GY}$. The curve AMB thus lies on a circle centered at Y. We leave the proof of the converse as an exercise.

Conic Sections

Figure 7 shows an arc AMB of a hyperbola with focus F and directrix $\mathcal{L}$. The points of the hyperbola are twice as far from F as from $\mathcal{L}$. To construct the normal at M, L'Hôpital generalizes Proposition X by taking some or all of the foci to be "straight or curved lines, which the straight lines MF, MG, etc. meet at right angles." Figure 7 illustrates his construction for the hyperbola $u - 2v = 0$, where $u = MF$, and $v = MG$ is the distance from M to the directrix. The normal MP passes through the center of gravity X of weights w_C and $w_{D'} = 2w_C$.

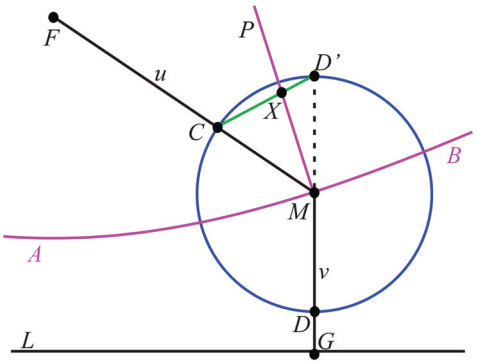


Figure 7 Constructing the normal to the hyperbola $u - 2v = 0$ at M.

Figure 8 shows the details behind the construction. To represent the differential dv, L'Hôpital drops the perpendicular mg to $\mathcal{L}$ from the point m of AMB infinitely close to M. He then drops the little perpendicular MS from M to mg, so that $dv = Sm$. Since $du - 2dv = 0$, the weight at D is negative, and L'Hôpital reflects D in M to D'. With a circle of arbitrary radius centered at M, he then constructs visible copies $\triangle MLC$ and $\triangle MKD'$ of the differential triangles $\triangle MRm$ and $\triangle MSm$. Writing the differential relation $du - 2dv = 0$ as $CL - 2D'K = 0$, L'Hôpital constructs the normal MP through the point X between C and D', with $CX : XD' = 2 : 1$.

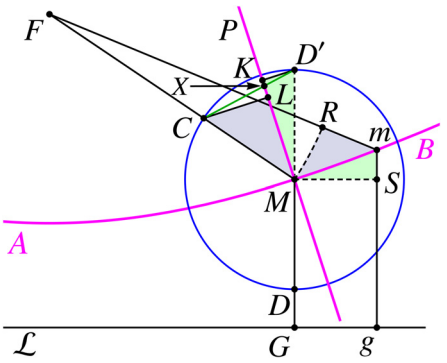


Figure 8 Constructing the normal to the hyperbola $u - 2v = 0$ at M.

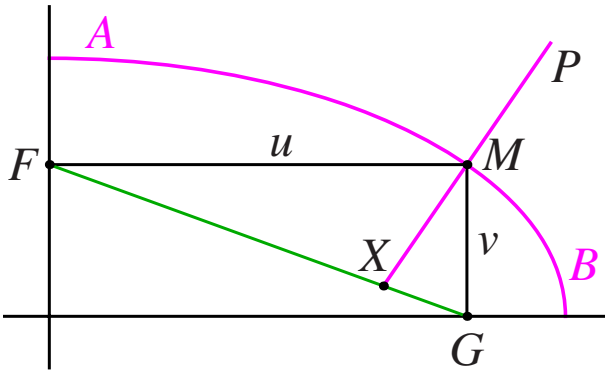


Figure 9 Constructing a normal to the ellipse $u^2/4 + v^2 = k$.

Figure 9 illustrates another example. The foci are a pair of perpendicular lines and the curve is the ellipse with equation

$$\frac{u^2}{a^2} + \frac{v^2}{b^2} = k.$$

We leave it as an exercise to show that the normal passes through the point X on FG, where $FX : XG = a^2 : b^2$.

Thread and Pen Constructions

For his last generalization, L'Hôpital takes some or all of the foci F, G, etc. to be "curved lines that have fixed and invariable origins at the points F, G, etc." and measures *wrap-around distances* from a point M to the foci.

For example, in Figure 10(a) the focus is an ellipse with origin F. For a point M outside the ellipse, the wrap-around distance $u = |MQ| + \ell(QF)$ is measured along the tangent segment MQ and the counterclockwise arc QF of the ellipse. With the same focus and origin, the wrap-around distance $v = |MR| + \ell(RF)$ is along the tangent MR and the clockwise arc RF. The curve AMB in Figure 10(b) has equation

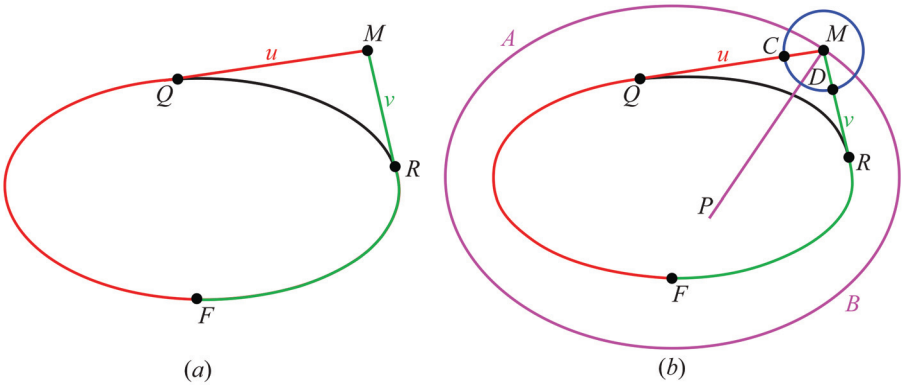


Figure 10 Wrap-around distances (a); a thread construction of an ellipse (b).

$u + v = k$, where k is a constant. Imagine wrapping a closed inelastic string of length k around the ellipse and pulling it taught with a pen at M. Move the pen and it traces the curve AMB.

We omit the details of the proof, but L'Hôpital's construction of the normal proceeds much as before. He draws a small circle about M that intersects the tangent segments at C and D. Then since $du + dv = 0$, the normal passes through the center of gravity of weights $w_C = w_D$. This shows the normal bisects $\angle QMR$.

When the elliptical focus of Figure 10 degenerates to a segment QR, we have $u + v = k - QR$, and the curve AMB is an ellipse with foci Q and R (Figure 11). The curve AMB of Figure 10(b) is also an ellipse, a result first discovered by Charles Graves [**2**].

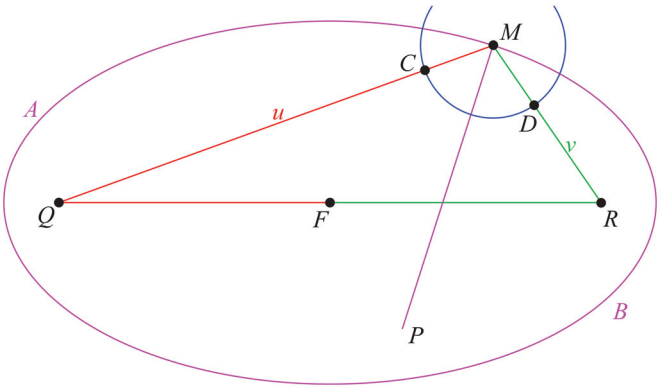


Figure 11 The standard thread construction of an ellipse.

Proposition X and the Chain Rule

To understand Proposition X in the context of the gradient vector and the chain rule, let's look at a few of the previous examples from a modern perspective. We begin with the Cassini oval and reestablish the rectangular coordinate system of Figure 1(a). The oval is a level curve of the function $f(x, y) = u(x, y)v(x, y)$, where

$$u = u(x, y) = \sqrt{(x + 1)^2 + y^2}$$

and

$$v = v(x, y) = \sqrt{(x - 1)^2 + y^2}$$

measure the respective distances from $M(x, y)$ to the foci $F(-1, 0)$ and $G(1, 0)$. The gradient vector

$$\nabla f = \nabla(uv) = v\nabla u + u\nabla v$$

gives the direction of the normal to the oval $f(x, y) = k$. Now the gradient

$$\nabla u = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) = \frac{(x + 1, y)}{\sqrt{(x + 1)^2 + y^2}}$$

is the unit vector parallel to $\overrightarrow{FM}$. This is not a surprise; at M, the function u increases at the fastest rate in the direction directly away from F. Furthermore, $|\nabla u| = 1$; taking a step of length $\triangle r$ directly away from F increases u by $\triangle r$. Similarly, ∇v is the unit vector parallel to $\overrightarrow{GM}$.

These observations lead to another interpretation of L'Hôpital's construction in Figure 3. Since $|\overrightarrow{MC}| = |\overrightarrow{MD}|$, the vector

$$(u + v)\,\overrightarrow{MX} = v\overrightarrow{MC} + u\overrightarrow{MD} = -|\overrightarrow{MC}|\,\nabla f$$

is anti-parallel to ∇f and hence normal to the oval.

A similar interpretation applies to the construction in Figure 7, where the hyperbola is a level curve of the function $f(x, y) = u(x, y) - 2v(x, y)$. Here ∇v is a unit vector that points directly away from $\mathcal{L}$ and

$$\begin{aligned} 3\overrightarrow{MX} = \overrightarrow{MC} + 2\overrightarrow{MD'} &= \overrightarrow{MC} - 2\overrightarrow{MD} \\ &= -|\overrightarrow{MC}|\,(\nabla u - 2\nabla v) = -|\overrightarrow{MC}|\nabla f. \end{aligned}$$

Similarly, for the wrap-around distance functions u and v of Figure 10(b), ∇u and ∇v are unit vectors pointing directly away from Q and R, respectively [**3**, p. 73]. So $\overrightarrow{MC} + \overrightarrow{MD}$ is parallel to the vector

$$-\nabla(u + v) = -(\nabla u + \nabla v)$$

and hence normal to the curve AMB.

These examples suggest a way to state Proposition X. We define a differentiable function $u : \mathcal{R} \mapsto (0, \infty)$ on an open subset $\mathcal{R} \subset \mathbb{R}^2$ to be a *distance function* if $|\nabla u(x, y)| = 1$ for all $(x, y) \in \mathcal{R}$. Now let $u, v : \mathcal{R} \mapsto (0, \infty)$ be distance functions and let $F : \mathcal{D} \mapsto \mathbb{R}$ be a function defined on an open set $\mathcal{D} \subset \mathbb{R}^2$. Suppose also that for each $(x, y) \in \mathcal{R}$, we have $(u(x, y), v(x, y)) \in \mathcal{D}$. Proposition X describes a way to construct the normal to the level curve

$$f(x, y) = F(u(x, y), v(x, y)) = k,\ k \in \mathbb{R} \text{ a constant}, \tag{3}$$

at any of its points in the xy-plane. While L'Hôpital allows for any function F, we should require that F be differentiable. Furthermore, to guarantee the level set (3) is a curve, we assume $\nabla F(u, v) \neq (0, 0)$ for each $(u, v) \in \mathcal{D}$.

L'Hôpital begins by taking differentials of the equation $F(u, v) = k$, giving

$$\frac{\partial F}{\partial u}\,du + \frac{\partial F}{\partial v}\,dv = 0.$$

Then, as in Figure 3, he constructs the normal through the center of gravity X of weights $w_C = \partial F/\partial u$ and $w_D = \partial F/\partial v$. With this notation, L'Hôpital's claim is that the normal is parallel to the vector

$$\mathbf{n} = \frac{\partial F}{\partial u}\overrightarrow{MC} + \frac{\partial F}{\partial v}\overrightarrow{MD} = (w_C + w_D)\,\overrightarrow{MX}. \tag{4}$$

Viewed from this perspective, Propostion X is an immediate consequence of the chain rule. For with $z = F(u(x, y), v(x, y))$,

$$\frac{\partial z}{\partial x} = \frac{\partial F}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial x}$$

and

$$\frac{\partial z}{\partial y} = \frac{\partial F}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial y}.$$

To give the chain rule a more geometric flavor, let us write these equations as

$$\begin{aligned}\nabla z = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) &= \frac{\partial F}{\partial u}\left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) + \frac{\partial F}{\partial v}\left(\frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}\right) \\ &= \frac{\partial F}{\partial u}\nabla u + \frac{\partial F}{\partial v}\nabla v,\end{aligned}$$

thus expressing the gradient ∇z as a linear combination of the gradients ∇u and ∇v. Note that the weights in the linear combination match L'Hôpital's weights w_C and w_D. Furthermore, when u and v are distance functions, $|\nabla u| = |\nabla v|$ and the vector $\mathbf{n}$ in equation (4) is anti-parallel to ∇z. Hence, L'Hôpital's line MP is indeed normal to the level curve (3).

We should keep in mind, however, that the gradient vector was first introduced by William Hamilton some 150 years *after* the publication of the *Analyse*. So we might think of L'Hôpital's Propostion X as a precursor to both the gradient vector and the chain rule. Furthermore, his arguments have a charm and geometric insight that the above computation seems to lack.

REFERENCES

[1] Bradley,R. E., Petrilli, S. J., Sandifer, C. E. (2015). *L'Hôpital's analyse des infiniments petits*. Switzerland: Birkhäuse.

[2] Poorrezaei, K. (2003). Two proofs of Graves's Theorem. *Amer. Math. Monthly* 110(9): 826-830. doi.org/10.2307/3647801

[3] Tabachnikov, S. (2005). *Geometry and Billiards*. Providence: American Mathematical Society.

Summary. In Chapter 2 of his *Analyse des infiniments petits* (*Analysis of the infinitely small*), L'Hôpital describes a method of drawing normals to a class of curves he calls generalized conics. We explore some applications and suggest how his Proposition X anticipates a chain rule for multivariable calculus.

FRED KUCZMARSKI (MR Author ID: 944646) received his Ph.D. in mathematics from the University of Washington. He enjoys baking bread and listening to nature in Washington's Cascades.

Infinite Series as Sums of Triangular Areas

HANS MUSGRAVE
University of North Dakota
Grand Forks, ND, 58202-8376
hans.musgrave@gmail.com

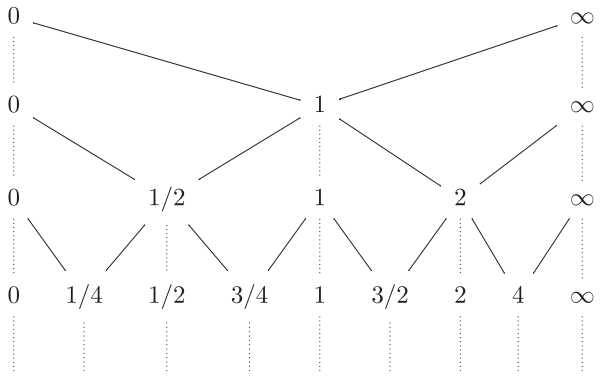
RYAN ZERR
University of North Dakota
Grand Forks, ND, 58202-8376
ryan.zerr@und.edu

The Riemann approach to integration associates the area under an integrable function's graph with rectangles. In this paper we take a different approach, exhausting the area using triangles. The methods are straightforward, involving only basic ideas from calculus and vector calculus, and the change in perspective is interesting in its own right. Because progress in mathematics often comes by reframing a problem, this particular example of a shifted perspective on a familiar topic may be instructive to anyone who is familiar with the varied historical approaches to integration.

Our approach relies on generating a dense subset of the interval of integration through the use of an infinite binary tree. This tree generates a series, and because the terms of the series arise from both the function defining the region and the choice for the nodes of the tree, it becomes possible to generate both familiar and exotic series and to know readily whether they diverge or converge and, if convergent, to what value. A particularly noteworthy example comes from using the Stern-Brocot tree, in which case connections arise with certain types of zeta functions. This connection provides a hint at the way the Riemann hypothesis is connected with topics such as Farey sequences.

A motivating example

To illustrate a few key ideas, consider the function $f(x) = 1/x$ on $(0, \infty)$. Imagine generating a countable dense subset of $(0, \infty)$ by starting with 0 and ∞ and placing between them the additional point 1. Between each of these three, generate two additional points: $1/2$ and 2. These five points are used to create four more by dividing the finite intervals into halves and adding the next power of 2 "between" the current largest finite value and ∞. The general process can be illustrated via an infinite tree:



Math. Mag. **95** (2022) 14–20. doi:10.1080/0025570X.2022.2001253

We now view the x and y-axes as the tangent lines to the graph of f "at" the points 0 and ∞, and we generate a new tangent line to the graph of f at the first point in the tree which lies between 0 and ∞. The resulting three tangent lines intersect, forming a triangle that lies below the graph of f, covering a portion of the area in the first quadrant. We repeat the process, except now forming triangles in the so-far uncovered area by considering the tangent lines at the next two points in the tree: $1/2$ and 2. This is shown in Figure 1.

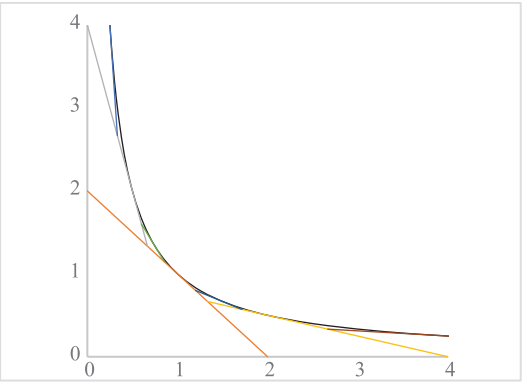


Figure 1 The first three tangent lines used in our triangular partition of the area under the curve.

In the limit, these triangles will exhaust the entire area under the curve. To further illustrate our point, if we consider the right-most or left-most triangles at each iteration, we see they all have an area of $2/3$—and thus can readily (and unsurprisingly) conclude that the area is infinite because the corresponding infinite series diverges. In this way we can establish a connection between the area under the curve and the infinite series corresponding to the combined area of the covering triangles. We investigate the nature of these series for different functions and different countable dense subsets.

Generalizing the approach

The standard approach to the Riemann integral amounts to utilizing a sequence of successively finer partitions to generate approximating areas whose limit is the area under the curve. Our approach is similar, except that the sequence of partitions must be regulated enough to ensure the approximating triangles can be conveniently defined. This is where the infinite binary tree comes in—it ensures partitions of sufficient regularity.

With such a sequence of partitions in hand, the process for generating the triangles which exhaust the area under the graph of a function f relies only on a few key properties:

1. The function must be defined on an interval of the form $(0, \infty)$, $[0, \infty)$, $(0, a]$, or $[0, a]$ for $0 < a < \infty$ and be tangent to the x- and y-axes at 0 and (if appropriate) a, respectively, or have one or both of the axes as asymptotes.
2. The function must be twice continuously differentiable, strictly monotonically decreasing, and strictly concave up, i.e., $f'(x) < 0$ and $f''(x) > 0$ for all $x \in A$ where A is the interval specified in Property 1.
3. The set generated by the tree must be a dense subset of A.

These three conditions are sufficient. Property 1 establishes that two of the legs of the first triangle will be the two axes, while Property 2 ensures that each newly-generated tangent line is distinct from those generated in previous steps of the process and thus yields the third side of a new triangle which does not overlap (except at the boundaries) any of the previous triangles. It also implies the triangle so formed lies below the curve. Finally, Property 3 means that every point (x_0, y_0) in the first quadrant and below the graph of the function will be in at least one triangle. In particular, by choosing a point on the tree sufficiently close to x_0 to ensure that the resulting tangent line lies above y_0 one ensures (x_0, y_0) is included in the covering.

Calculating the area of any one of these triangles is an application of basic calculus. For instance, if $0 < x_1 < x_{1,2} < x_2$ are three consecutive points at a given level of the tree, with $x_{1,2}$ having been generated from the two previously consecutive points x_1 and x_2, then one can find the tangent lines to the graph of f at each of x_1, $x_{1,2}$ and x_2, use them to find the corresponding three vertices of the resulting triangle, and—by finding the cross product of the vectors describing two of the triangle's sides—obtain an area formula of the form Area $= A^2/2B$ where

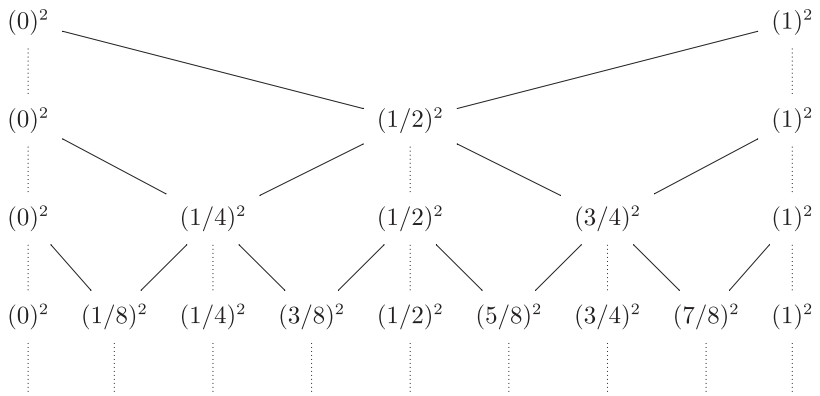
$$\begin{aligned} A = {} & f'(x_1)f'(x_{1,2})(x_1 - x_{1,2}) + f'(x_1)f'(x_2)(x_2 - x_1) \\ & + f'(x_2)f'(x_{1,2})(x_{1,2} - x_2) + f(x_1)(f'(x_2) - f'(x_{1,2})) \\ & + f(x_2)(f'(x_{1,2}) - f'(x_1)) + f(x_{1,2})(f'(x_1) - f'(x_2)) \end{aligned}$$

and

$$B = (f'(x_1) - f'(x_{1,2}))(f'(x_1) - f'(x_2))(f'(x_2) - f'(x_{1,2})).$$

An illustrative example

Consider the function $g(x) = (1 - \sqrt{x})^2$ on $[0, 1]$ and the tree



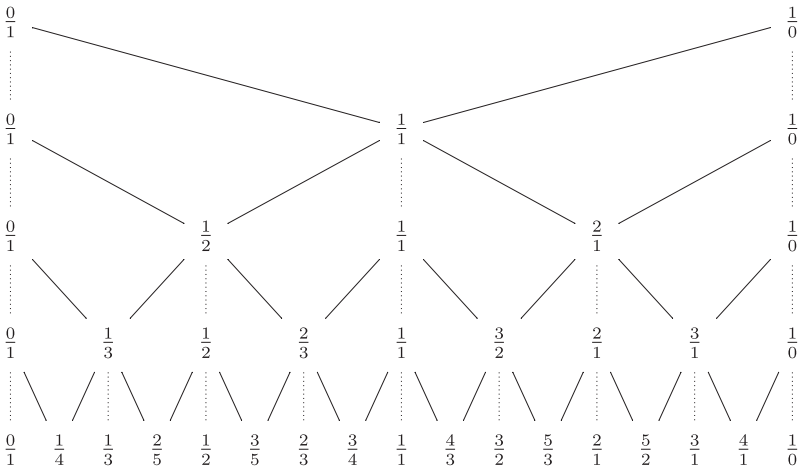
which represents a modified version of the tree presented above. This situation meets the three properties specified earlier, and we find that if we consider level n of this tree, letting k range between 0 and $2^n - 1$, it is possible to calculate the area of each triangle generated at the n-th tree level. These areas turn out to be independent of k, all equaling $1/2^{3n+3}$. Thus, for each $n \geq 0$ and each $0 \leq k \leq 2^n - 1$ the area of the corresponding triangle is $1/2^{3n+3}$. Thus, the area under the graph of $g(x)$ is given by

$$\sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \frac{1}{2^{3n+3}} = \sum_{n=0}^{\infty} 2^n \cdot \frac{1}{2^{3n+3}} = \frac{1}{6}.$$

It should be noted that this is in no way a surprising conclusion, as the area may also be calculated by evaluating $\int_0^1 g(x)\,dx$ directly. However, this example illustrates the variety of ways one may conveniently generate the dense subset A of the domain of the function. Modifications such as this can be useful for simplifying the form of the terms in the corresponding series.

Generating dense subsets using the Stern-Brocot tree

The example of $g(x)$ on [0, 1] becomes more profound when used in conjunction with a dense subset generated by the Stern-Brocot tree:



The most general form of the tree is generated by starting with the rationals $0/1$ and $1/0$ and generating successive vertices by forming the *mediant* of those adjacent rationals already formed as part of the iterative process. The mediant is simply found by separately adding numerators and denominators of the two given rationals. It turns out (see, e.g., Graham, Knuth, and Patashnik [2]) that if m/n and m'/n' are adjacent rationals at some level in the Stern-Brocot tree, then $mn' - m'n = -1$. If we use this fact, focus on the left half of the Stern-Brocot tree (that generated by starting with $0/1$ and $1/1$), and apply the area formula given above to $g(x)$, the area may be simplified to

$$\frac{1}{2n^2(n')^2(n+n')^2}$$

provided we take the dense subset A of [0, 1] in this case to be generated by squaring the value at each vertex of the tree. Hence, given our conclusion about the total area from the previous section, we arrive at

$$\sum_{l=0}^{\infty} \sum_{\forall n_l} \frac{1}{2n_l^2(n_l')^2(n_l+n_l')^2} = \frac{1}{6},$$

where n_l and n_l' are adjacent denominators at level l of the tree. Thus, we sum over all levels and all pairs of adjacent denominators in each level.

Note how changing the nature of the points constituting the subset A leads to a different form for the terms of the infinite series. In fact, this last result, with additional details, can also be found in Kramer and von Pippich [3], where it is noted that

this series is a special instance of a more general type of function called a Mordell-Tornheim Zeta Function,

$$\sum_{(m,n)=1}^{\infty} \frac{1}{m^{s_1} n^{s_2} (m+n)^{s_3}}.$$

Remarkably, related ideas are connected, via Farey Sequences, to the Riemann Zeta Function and the Riemann Hypothesis [1, 4, 5].

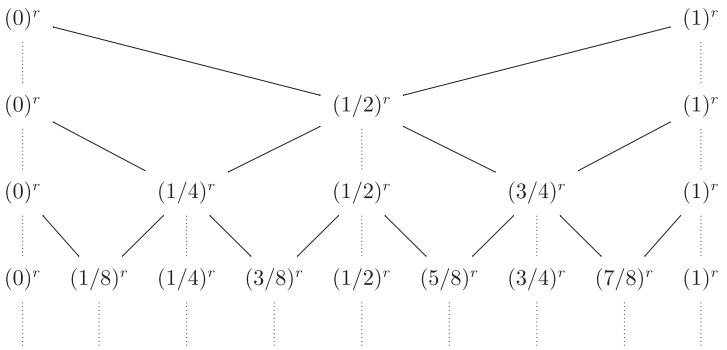
Thus, via the same function $g(x)$ and domain $[0, 1]$, we obtain—by altering the process through which the points of A are generated—a much different, much less easily evaluated, convergent series.

A class of convergent series

Applying the above ideas to $y = 1/x$ using either the tree based on powers of 2 or the Stern-Brocot tree yields a simple formula for the corresponding divergent series, although in neither case is the conclusion of divergence surprising given the form of the series' terms – which allow for the conclusion of divergence in a more direct way. The reader may also reasonably wonder about situations in which the portion of the unit circle $(x-1)^2 + (y-1)^2 = 1$ for $0 \le x, y \le 1$ or a function like $y = (1 - \sqrt[3]{x})^3$ are used on the interval $[0, 1]$ to generate convergent series. In both cases straightforward modifications of the terms constituting the tree make helpful simplifications to the calculations necessary to determine the areas of the covering triangles. In neither case, however, is the result a series with an "elegant" or simple form.

On the other hand, taking r to be a natural number greater than 1 and forming the function $f(x) = x^{-1/r}$ on $[0, 1]$ does allow for a conveniently expressible form for the resulting family of infinite series, and illustrates one additional generalization to the ideas outlined in this paper.

Focusing first on the tree:



it is possible to conclude a general triangular area is given by the fraction A/B, where

$$A = (1+r)^2 \left[(k+1)^r (2k+1)^r + k^r (2k+1)^r - 2^{r+1} k^r (k+1)^r\right]^2$$

$$B = r 2^{nr-n+1} \left[(2k+2)^{r+1} - (2k+1)^{r+1}\right] \times \left[(2k+1)^{r+1} - (2k)^{r+1}\right] \left[(k+1)^{r+1} - k^{r+1}\right],$$

and thus, for instance, with $r = 2$ we have the series

$$\sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \frac{9(6k^2+6k+1)^2}{2^{n+2}(3k^2+3k+1)(12k^2+6k+1)(12k^2+18k+7)}.$$

No doubt the reader has noted that this example violates one part of the three properties outlined earlier in the paper—namely that $f(x)$ be tangent to both axes on $[0, 1]$. Thus, it is not possible to conclude directly that the above series converges to the value of $\int_0^1 f(x)\,dx$. However, the necessary adjustment is straightforward and easily discerned by examining Figure 2, showing which portion of the area under the graph of f is not covered by triangular areas. In particular, for the case where $r = 2$ we see that

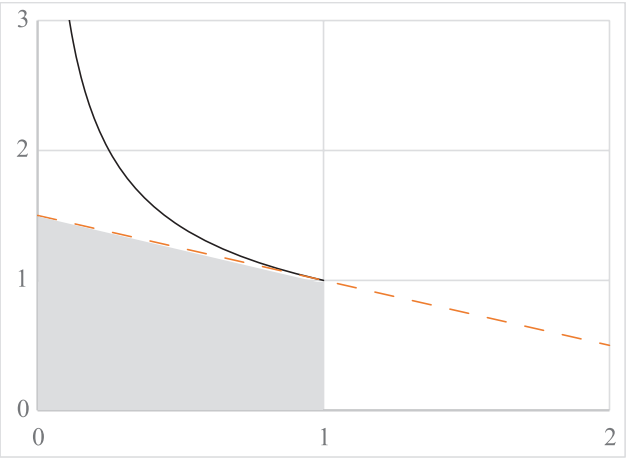


Figure 2 The region under the graph of f not covered by triangular areas.

$$\sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} \frac{9(6k^2+6k+1)^2}{2^{n+2}(3k^2+3k+1)(12k^2+6k+1)(12k^2+18k+7)}.$$

converges to the value of

$$\int_0^1 \frac{1}{x^{1/2}}\,dx - \frac{5}{4} = \frac{3}{4},$$

where $5/4$ is the area of the shaded trapezoid. More generally, if we consider the triangular area corresponding to $f(x) = x^{-1/r}$, the series whose terms are the fractions A/B given above converges to

$$\int_0^1 \frac{1}{x^{1/r}}\,dx - \frac{2r+1}{2r} = \frac{r+1}{2r(r-1)}.$$

Summary

Rather than the standard approach using rectangles, we instead approximate the area under the graphs of certain functions using triangles generated by well-regulated partitions which arise from binary trees. The series which arise have forms that depend on the structure of the tree, and examples are given illustrating how markedly different series turn out to correspond to the same definite integral. The interested reader may want to consider things from the opposite direction; namely, given a series, under what circumstances will there exist a function for which the techniques of this paper lead to that series?

REFERENCES

[1] Franel, J. (1924). Les suites de Farey et le problème des nombres premiers. *Nach. von der Gesellschaft der Wissenschaften zu Gött., Math.-Phys. Klasse*, 1924: 198-201.
[2] Graham, R. L., Knuth, D. E., Patashnik, O. (1989). *Concrete Mathematics*. Reading: Addison-Wesley.
[3] Kramer, J., von Pippich, A-M. (2016). Snapshots of modern mathematics from Oberwolfach: Special values of zeta functions and areas of triangles. *Notices of the American Mathematical Society* 63(8): 917-922.
[4] Passare, M. (2008). How to compute $\sum 1/n^2$ by solving triangles. *Amer. Math. Monthly*. 115(8): 745-752. doi.org/10.1080/00029890.2008.11920587
[5] Tou, E. R. (2017). The Farey sequence: From fractions to fractals. *Math Horizons*. 24(3): 8-11. doi.org/10.4169/mathhorizons.24.3.8

Summary. A method is developed for exhausting the area under certain curves using triangles generated by an infinite binary tree. The associated series depend on the choice of binary tree, and examples are given to illustrate how this method can be used to evaluate certain series.

HANS MUSGRAVE received his M. S. degree in Mathematics from the University of North Dakota. He subsequently took a position as a data engineer with Dynamic Signal in Silicon Valley. His mathematical interests range from the present paper to optimization and non-standard analysis.

RYAN ZERR (MR Author ID: 758134) is a professor of mathematics at the University of North Dakota. His interests include dynamical systems, and it was work on a problem in this area that unexpectedly led to the present article.

Snakes: Legal, Illegal and Dodecahedral

THERESE AGLIALORO
University of Dallas
Irving, TX 75062
maglialoro@udallas.edu

ROBERT HOCHBERG
University of Dallas
Irving, TX 75062
hochberg@udallas.edu

Solved Rubik's cubes are all alike, but every unsolved cube is unsolved in its own way. Among these unsolved arrangements lies a wonderful array of beautiful patterns, and discovering them can be the occupation of many an enjoyable hour. One of the oldest of these patterns is the *snake*, which is formed by putting an "L" on each face so that they connect to form a contiguous shape winding around the cube, entering and leaving each face once. The authors wondered whether the Megaminx puzzle (Shown in Figure 1*) also had snakes, and if so, how many. Finding the answer required a pleasant tour of graphs, groups and geometry, which this article aims to share. And *our* answer to *how many* is 21, but arguments could be made for 6, or 700.

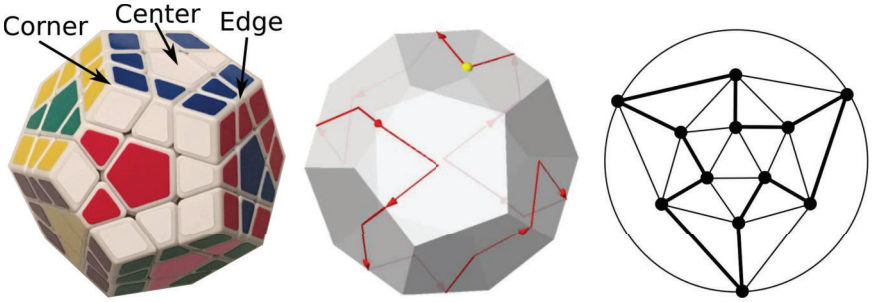


Figure 1 The Megaminx with a snake (left), a schematic view of the snake (middle), and the corresponding graph (right).

The Megaminx puzzle and its snakes

The Megaminx is a dodecahedral variant of the Rubik's Cube, shown on the left in Figure 1. It has twelve turnable faces, each of which contains one center piece, five corner pieces, and five edges pieces, as labeled in the figure. Center pieces have one pentagon-shaped sticker, corner pieces contain three parallelogram-shaped stickers, and edge pieces contain two trapezoid-shaped stickers. The entire Megaminx puzzle has 12 center pieces, 20 corner pieces and 30 edge pieces. The Megaminx is considered "solved" when the eleven stickers on each face are all the same color. The Megaminx that we will consider has a different color for each solved face, so twelve colors altogether.

Math. Mag. **95** (2022) 21–31. doi:10.1080/0025570X.2022.2001286
MSC: 00A08, 05A15, 05C30, 52B15
*The online version of this article includes color diagrams.
Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/0025570X.2022.2001286

A *snake* on the Megaminx looks like a solved Megaminx, except that the twelve centers seem to have been moved, along with twelve edges, so that the moved pieces form one long cycle around the puzzle that enters and leaves each face exactly once. Such a snake is shown on the left in Figure 1, and schematically in the center of that figure. The figure on the right is a graph representation of that same snake. Graphs will be discussed in the section on corner-axis snakes. (Note that the term "snake" is sometimes used to refer to a chordless cycle in a graph, but we never employ that usage in this article.)

The two important aspects of a snake are *shape* (its geometric shape on the dodecahedron) and *realization* (how that shape is embedded onto a colored Megaminx puzzle, and how the pieces move to form that shape). Most of this article is concerned with shape. Only one paragraph in the last section talks about realization and color.

The two-step snake

We imagine building snakes on the Megaminx as a two-step process, starting from a solved puzzle. In Step 1 we move the twelve centers (see below), and in Step 2 we move twelve of the edges in such a way that those edges connect the twelve (already-moved) centers into a single circuit. This is probably not how a skilled puzzler would actually put a snake on the Megaminx, but it is a good way for us to carry out our analysis. Let us consider Step 1.

The Megaminx is constructed around a single central mechanism, shown on the left in Figure 2, that connects all twelve centers together. These center pieces directly hold in 30 edge pieces, and together the centers and edges hold in the corner pieces. The image on the right of Figure 2 shows the Megaminx with just twelve edges put in, forming the snake shown earlier. The Megaminx is manipulated by turning faces, which rotates the center pieces, but does not change their locations. In this sense, it is not strictly correct to talk about "moving centers," even though that is how we think about Step 1. To "move the centers," one essentially rotates the entire Megaminx in space and then carries out face turns to move the edge and corner pieces back to their starting positions.

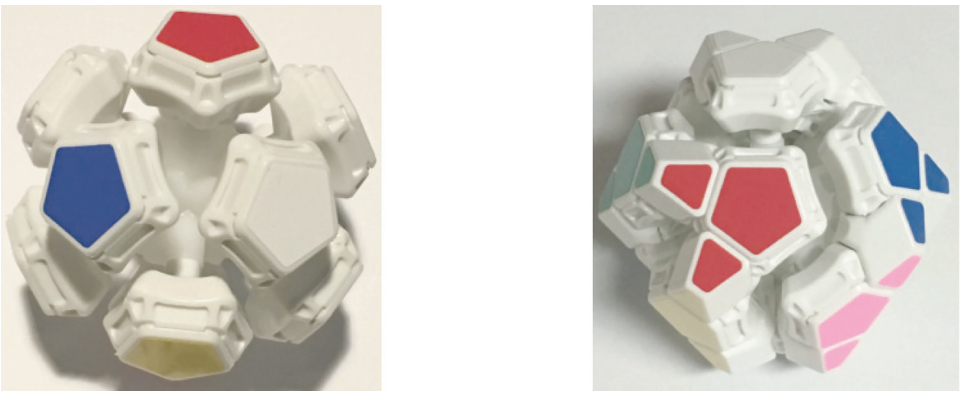


Figure 2 (On the left, the central mechanism of the Megaminx puzzle. On the right, we have enough edges to make a snake.

Imagine that we remove the central mechanism (somehow leaving corners and edges intact), move it around in space, and then put it back. This can be done in 60

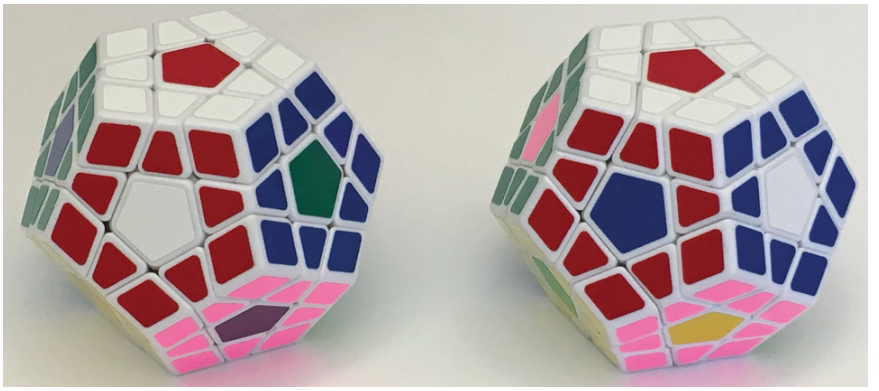


Figure 3 All twelve centers moved: 180° about the red/white edge axis (left), 120° clockwise about the red/white/blue corner axis (right). These patterns may be obtained from a solved Megaminx via ordinary face turns.

ways: there are twelve different colors that could end up on top, and five ways the mechanism can be rotated once a top color has been chosen. A theorem of Euler (see Palais, Palais, and Rodi [7] for a thorough treatment) states that every symmetry of a 3D solid corresponds to some rotation about an axis. Table 1 summarizes these rotations. We include a row for the identity symmetry (which leaves the central mechanism unmoved), thereby accounting for all 60 symmetries. The "# Axes" column counts the number of axes of the given type. For example, there are 12 faces. A face axis passes through the centers of two opposite faces, giving $\frac{12}{2} = 6$ face axes. The "Rotations Per Axis" column counts how many different rotations are possible about each given axis. For example, if we imagine an axis passing through the center of the white face, as well as its opposite face, then 72°, 144°, 216°, and 288° clockwise rotations are possible. The last column gives the total number of symmetries of each type, and is simply the product of the number of axes of that type and the number of rotations per axis. We note that this column sums to 60, so all symmetries are accounted for.

Type of axis	# Axes	Rotations per axis	# Symmetries
Identity	—	—	1
Face Center	6	4	24
Edge Center	15	1	15
Corner	10	2	20

TABLE 1: The 60 symmetries of the dodecahedron

Step 1 requires that no face has the same center with which it started after the centers have been moved, meaning that the Identity and Face Center symmetries are unsuitable for snake construction. The other two types of symmetries do move all twelve centers, and it is possible to perform these last two types of central mechanism symmetries by legal moves starting from a solved puzzle, as shown in Figure 3. We consider these two symmetries in turn.

Corner-axis snakes

In the case of corner-axis snakes, the centers have been rotated 120° about an axis passing through two opposite corners. We want to model the motion of 12 edges that will connect these 12 moved centers into a cycle. We use a graph to model the puzzle, shown on the right in Figure 1. It contains one vertex for each face of the Megaminx, and edges connecting vertices corresponding to faces of the Megaminx that have an edge piece in common. Thus, each vertex is connected to exactly five other vertices because each face of the dodecahedron shares an edge with exactly five other faces. We note that this graph is usually considered the *icosahedron* graph, which is the *dual* of the dodecahedron. This graph is more suitable to our purposes because we want our vertices to correspond to faces of the Megaminx. (See West [**9**, Ch. 7], or Wilson [**10**, Ch. 3] for more on graphs and duals.)

The center of that drawing corresponds to one corner through which the axis passes, and the three vertices forming the triangle in the center correspond to the three faces that meet at that corner. The graph has 120°-rotational symmetry, corresponding to the 120° rotation of the central mechanism. If we highlight the twelve moved edges on this graph, then we should get a single circuit of edges that hits each vertex exactly once. Such a circuit is called a *Hamilton circuit*, named in honor of William Rowan Hamilton (who, in a pleasing connection to the present problem, first made them famous by marketing a dodecahedral puzzle called the Icosian Game). Furthermore, this circuit must itself be 120°-rotationally symmetric because any edge that moves must move with its adjacent centers, so that the sticker colors will match. Let us find all such Hamilton circuits, if any exist.

We first observe that edge a in Figure 4 cannot be in a snake. If it was, then b and c would also have to be in that snake, but those three edges would form a triangle, which cannot exist in a single circuit containing 12 edges. For the same reason, edges b and c cannot be selected. A snake cannot contain both edges d and e, for they would form, with their symmetric edges, a circuit of length 6. Since every vertex must meet two edges of the circuit, and considering the vertex between edges d and e, we conclude that edge f must be used, as well as exactly one of d or e. Without loss of generality, assume that edge e is used. By similar reasoning, edge g must be used. Highlighting these edges, and those symmetric to them through 120° rotations, we obtain the second diagram of Figure 4. There is only one way to complete a Hamilton circuit now, and the result is shown in the third diagram of Figure 4. This gives us a candidate shape for a snake on the Megaminx, but is it actually achievable? Indeed it is! It is the one shown in Figure 1, which was obtained from a solved puzzle by making ordinary face turns.

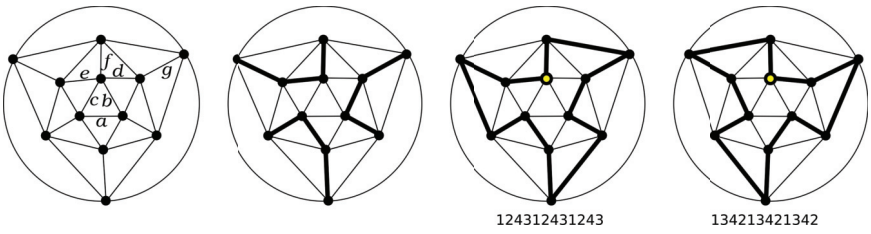


Figure 4 Deriving a snake with 120°-rotational symmetry.

In the previous paragraph we selected edge e. If we had selected d instead, then the result would have been the mirror image of the one just found, as shown on the right

in Figure 4. They are also mirror images when rendered on the Megaminx, as snakes. However, these snakes are even more alike than their graphs suggest. To see how, we introduce the *turn sequence* of a snake.

Turn sequences

The turn sequence of a snake completely describes its geometric shape and is obtained as follows: The snake enters each face across some edge and departs along a different edge. The *turn* at that face is either 1, 2, 3, or 4, depending on how many steps counter-clockwise the departing edge is from the entering edge. The *turn sequence* is obtained by following the snake on the graph (or dodecahedron) and writing down the twelve turns. Figure 4 shows the turn sequences for our first two snakes, starting from the light-colored vertex and proceeding counter-clockwise.

If we follow some Hamilton circuit in the opposite direction (but still measure each turn counter-clockwise at each face), then we visit the vertices in reverse order, and the turns that we observe at each face are five minus the original turn value, which we will call the *complement*. That is, $\text{complement}(x) = 5 - x$. We call the resulting sequence the *reverse complement* of the original sequence. Observing that the two turn sequences shown in Figure 4 are reverse complements of each other (suitably shifted), we conclude that the two snakes are not only mirror images of each other, but are actually the same snake, just followed in opposite directions. The two graphs show the (one) snake as seen from the two opposite poles of its axis of symmetry. In summary, there is, up to rotation, exactly one shape for a snake on the Megaminx with 120°-rotational symmetry.

Some observations about turn sequences: They are cyclic, meaning that we can read a turn sequence starting at any position as long as we wrap around at the end back to the beginning, and the resulting sequence will describe the same geometric snake. If we reverse a turn sequence, then the resulting snake will be a mirror image of the original. Unexpectedly (for us) the sum of the twelve turns will always be exactly 30. This is because the Hamilton circuit may be thought of as a polygon with twelve sides, and the edges inside that polygon constitute a triangulation of that polygon. Any triangulation of an n-gon requires $n - 3$ chords (see the wonderful article by Meister "Polygons have ears" [**5**]), so there will be exactly nine edges inside the polygon, and the remaining nine edges will be outside the polygon. The turn sequence counts the nine edges outside the circuit twice each, and each edge of the circuit once, yielding a sum of $2 \cdot 9 + 12 = 30$.

Legal edge-axis snakes

Now we consider whether snakes may be built on the 180° rotation of the central mechanism shown on the left in Figure 3. The axis of symmetry passes through the center of an edge, and so we draw our graph differently, as shown on the left in Figure 5, which has an edge center in the center of the drawing. The other edge through which this axis passes is the one shown in two pieces, going off the top of the graph and "wrapping around" to the bottom. We call these two edges *axis edges*. For the same reason as earlier, any Hamilton circuit corresponding to a snake must be 180° rotationally symmetric, and the middle and right diagrams in Figure 5 show two such circuits. The middle one does not use the axis edges, but the right one does.

First, we find all Hamilton circuits that do not use the axis edges. There are sixteen of them, as shown in Figure 6, grouped according to the two edges that meet the

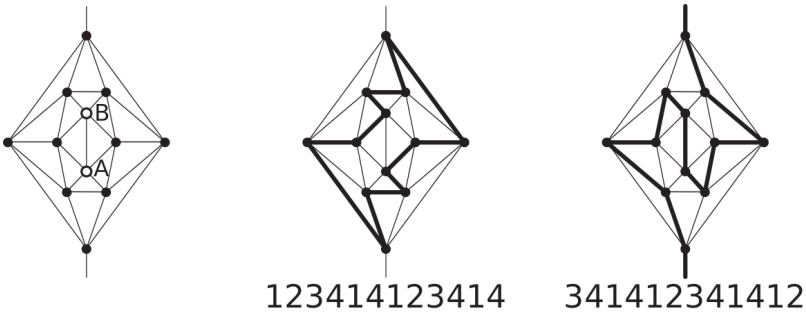


Figure 5 Deriving snakes with 180° symmetry. Graph drawn with edge symmetry (left). Hamilton circuit yielding a legal snake (middle). Hamilton circuit yielding an illegal snake, as it uses the axis edges (right). Turn sequences begin at vertex A.

bottom vertex. Note that there is no group that corresponds to selecting the left-most and right-most edges (can you see why?). Each of these circuits has been labeled with half of its turn sequence, namely the portion counter-clockwise from vertex A up to but not including vertex B. By symmetry, the second half of the sequence will be a repeat of the first half. Twelve of these circuits are reverse complements of others in the list, giving six pairs that really describe the same geometric snake. For example, the reverse complement of 141342 (A2 in Figure 6) is 312414 (D2) suitably shifted. Four of the snakes are their own reverse complements (shown in bold). Thus, there are exactly $\frac{12}{2} + 4 = 10$ snakes with 180° symmetry not using the axis edges. All of these may be obtained on the Megaminx by ordinary face turns, starting from a solved puzzle. We know, because we did them all!

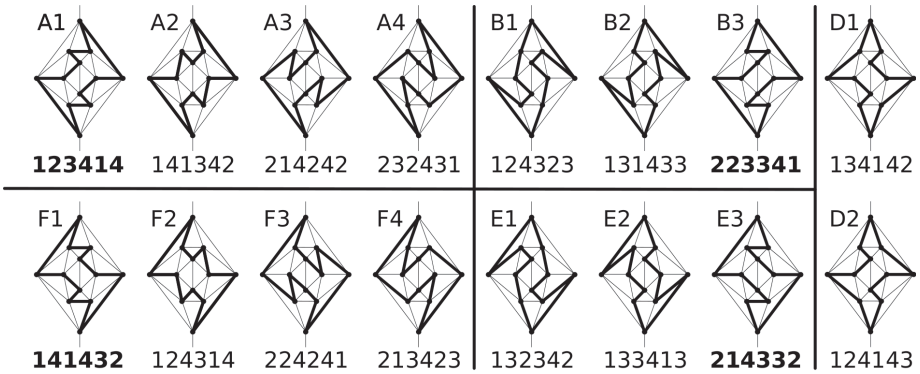


Figure 6 All legal Hamilton circuits with 180° symmetry about the central edge. Snakes in the same column are mirror images of each other.

Illegal edge-axis snakes

Now we consider the 180° symmetric snakes that do use the axis edges. They are given in Figure 7, grouped by the edge met by the bottom vertex, in addition to the "wrap-around" edge that all of these snakes use. The sequences begin at the bottom vertex,

122334	213414	312424	412334
122341	213421	312431	**412341**
124312	**214143**	**334122**	414312
131342	**221433**	**341412**	421342
134213	243124	342134	424213
141243	243131	342141	431243
143214			433214
143221			**433221**

Figure 7 All illegal Hamilton circuits with 180°symmetry about the central edge. The second half of each sequence is the reverse complement of the given half.

entered via the wrap-around edge. We leave it to the reader to draw them if desired. A printable template is available in the supplementary resources.

We give only the first half of each turn sequence because the turn sequences of these snakes are their own reverse complements. To see this, consider, for example, the circuit shown on the right in Figure 5. If we traverse the circuit in the order $A, B, \ldots,$ (using the labels in the left figure), then the turn sequence will be the same as that obtained by traversing the circuit in the order $B, A \ldots,$ by symmetry, but it will also generate the reverse complement, as discussed earlier. Also, if we shift any of these sequences by six places (and wrapping around), then we obtain another sequence in the list, namely, the same snake as seen from the opposite axis edge. This implies that our snakes come in pairs, except when a turn sequence is already equal to its own six-shift, which is equivalent to saying that its first half is already its own reverse complement. This is the case for the eight bold sequences in Figure 7. Since the remaining 20 come in pairs, we conclude that there are $\frac{20}{2} + 8 = 18$ shapes for these snakes.

Now, here something interesting happens. Snakes with these shapes *cannot* be realized on the Megaminx by ordinary face turns starting from a solved puzzle. The reason has to do with the number of swapped pairs of edges. We are moving twelve edges, but two of them—the axis edges—merely rotate in place, but do not change their locations. The snake's remaining ten edges consist of five pairs that swap locations with each other, and "five swaps" corresponds to an odd-parity permutation of the edges. It can be shown, using a small bit of group theory, that odd-parity permutations of edges are simply unachievable from a solved puzzle by ordinary face turns. For a very entertaining treatment of groups, permutations, parities, and puzzles, see Kiltinen [**4**]. Muralidharan [**6**] applied these ideas to the 15-puzzle. Ewing and Kośniowski [**3**] and Frey and Singmaster [**8**] applied them to the Rubik's cube. And Bandelow [**1**] and Eidswick [**2**] applied them to the Megaminx, there called the"magic dodecahedron" and "Halpern's dodecahedron," respectively.

However, let us think outside the subgroup (a funny joke if you have looked at the above references). If we start with a solved Megaminx, remove the white-red edge piece (that is, the edge piece that has white and red stickers), put it where the white-blue edge piece goes, and put the white-blue edge piece where the white-red edge piece goes, ensuring that the white face is still solid white, then we have put an odd parity onto the puzzle, and all those axis-edge snakes now become reachable. Again, we did them all! Since these snakes can be achieved only after swapping two edge pieces, we call them "illegal."

So how many snakes are there?

We begin by counting the total number of Hamilton circuits of our graph. We wrote a Python program (available with the online resources) to enumerate these, and found that there are 1280 of them. (Note that some sources cite double that number, counting each circuit twice since it could be traversed in two different directions.) When we group together those that are geometrically identical to each other (as indicated by their turn sequences) we find 33 of them. This is shown in Table 2, where they are grouped according to their symmetries. The "#Sym." column gives the number of symmetries of each shape, including the identity symmetry. The "#Circ." column gives the number of Hamilton circuits on the dodecahedron that have this shape; this entry is equal to 60 divided by the number of symmetries. The left turn sequence in each row is the mirror image of the right turn sequence in that row. Recall that the 120° snake sequence is its own mirror image, upside-down, so it has no pair.

If you count only the shapes of legal snakes, and consider mirror images equal to each other, then there are only the six given in the top three groups of that table. If we do not identify mirror images, then there are eleven shapes of legal snakes. If we also allow illegal snakes, then we bring in the fourth row of the table, and those numbers become eleven and 21, the latter being our favorite answer.

The turn sequences in the bottom row of the table cannot be realized as snakes on the Megaminx. This is because the rotational symmetry of any snake yields a symmetry in the turn sequence (shift by four for 120° rotations, shift by six for legal 180° rotations, and reverse complement for illegal 180° rotations). Sequences in the bottom row have no symmetry.

On the other hand, some Hamilton circuits on the dodecahedron can represent more than one snake. Here we consider realization and color, the second aspect of snakes. Those with 120° symmetry have two snakes each, as they can be rotated by 120° and 240°. Even more interesting are those in the top row of Table 2, the ones we call *hyperpossible*. You might have noticed that the sequence on the top-left of that row appears as a legal snake in Figure 6 (circuit B3) and also *twice* as an illegal snake in Figure 7 (the top bold entry in columns one and three). The same is true for the other sequences in that row. Hyperpossible circuits yield three distinct snakes!

Now we can see where 700 comes from, if we take the 12 colors of the Megaminx faces into consideration. Using the organization of Table 2, we count the number of snakes for each type of turn sequence. For example, each turn sequence in the top row gives one legal snake, which can be put on the Megaminx in 15 different ways. Since there are four circuits of that type, the row yields 60 legal colored snakes. That same row yields 120 illegal snakes, since each turn sequence yields two illegal snakes. Table 3 summarizes this count, giving 280 legal colored snakes, 420 illegal colored snakes, and 700 colored snakes total.

What about the cube?

Table 4 gives a summary of the foregoing analysis applied to the Rubik's cube. One observation worth mentioning is that the 180° motion of the centers in Step 1 is *illegal*—it can be be achieved only after swapping two edges—and if the axis of rotation does not pass through an edge of the circuit, then the snake requires three swaps of pairs of edges, making the resulting snake *legal*. Two wrongs have made a right. Also,

Turn sequences		#Sym.	#Circ.	Notes
122334122334 141432141432	143322143322 123414123414	4	15	These circuits have three orthogonal 180° axes of symmetry. They can be put on the Megaminx in one legal way, and two illegal ways
124312431243		3	20	The only 120° snake
122424122424 124143124143 124323124323	131433131433 124314124314 123243123243	2	30	These circuits have a single 180° axis of symmetry *not* passing through an edge of the circuit, and so correspond to legal snakes
122341412334 123421414143 124134213243 123421341243 131424213243	141432214332 124321341414 124314213423 124321342143 124241313423	2	30	These circuits have a single 180° axis of symmetry passing through an edge of the circuit, and so correspond to illegal snakes
122342141334 122431412424 122342214234 122432141424 123421421433 124321414314	124322143314 124143131433 122432214324 123414131433 122432134134 123414124143	1	60	These circuits have no symmetries, and so have no corresponding snakes

TABLE 2: All Hamilton circuits on the faces of a dodecahedron

Turn sequences	#Seqs.	#Circ.	#Legal	#Illegal
Hyperpossible	4	15	60	120
120°	1	20	40	0
180°Legal	6	30	180	0
180°Illegal	10	30	0	300
Impossible	12	60	0	0
Total			**280**	**420**

TABLE 3: There are 700 colored snakes on the Megaminx Puzzle

in addition to the snake shape mentioned in this paper's opening paragraph, there is another shape for these snakes, namely those that have a "2" in their turn sequence, creating an "I" instead of an "L" on some faces. Altogether, there are three different snake shapes, and 44 colored snakes, on the Rubik's cube.

A closing challenge

We close with a challenge for the reader. Every Megaminx snake, and indeed every Hamilton circuit on our graph, can be drawn as a curve on the surface of the dodecahedron similar to that shown in the middle of Figure 1, and that curve separates the

Turns	#Syms.	#Circ.	Notes	#Legal	#Illegal
131313	6	4	Has one 120° corner-axis and three 180° edge-axis symmetries. Each circuit gives rise to two legal and three illegal snakes.	8	12
123123 132132	4	6	Has three orthogonal axes of symmetry, one through faces and two through edges. Each circuit gives rise to one legal and one illegal circuit.	12	12

TABLE 4: Summary of snakes on the Rubik's cube

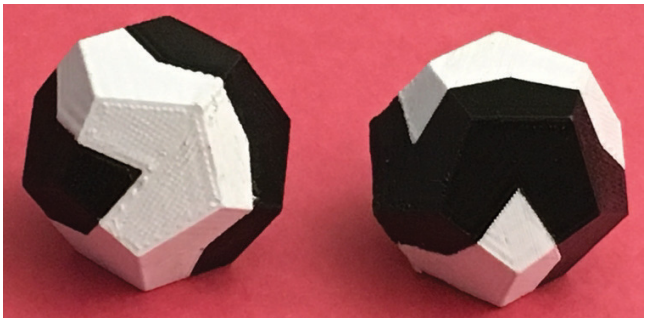


Figure 8 Every snake path divides the dodecahedron's surface into two regions with the same area.

surface into two regions. Figure 8 shows two models we constructed using a two-color 3D printer, where one region was printed in black, and the other in white. One line in this article implicitly contains a proof that those regions will always have the same area. (More practically, because of how our models are generated, they will each use the same amount of each color plastic.) We invite the reader to find that proof. We note that the result is true for all circuits, whether legal, illegal or impossible, and also holds on the other four Platonic solids as well. The STL files (which can be sent to 3D printers) for all of our snakes, in 1- and 2-color prints, are available with the supplementary materials.

REFERENCES

[1] Bandelow, C. (1982). *Inside Rubiks Cube and Beyond*. Boston, MA: Birkhäuser.

[2] Eidswick, J. A. (1986). Cubelike puzzles—what are they and how do you solve them? Amer. Math. Monthly. 93(3): 157-176. doi.org/10.1080/00029890.1986.11971778

[3] Ewing, J., Kośniowski, C. (1982). *Puzzle it out: Cubes, groups and puzzles*. Cambridge: Cambridge University Press.

[4] Kiltinen, J. O. (2003). *Oval Track and Other Permutation Puzzles, and Just Enough Group Theory to Solve Them*. Washington, DC: Mathematical Association of America.

[5] Meisters, G. H. (1975). Polygons have ears. *Amer. Math. Monthly*. 82(6): 648–651. doi.org/10.1080/00029890.1975.11993898

[6] Muralidharan, S. (2017). The fifteen puzzle—a new approach. *Math. Mag.* 90(1): 48-57. doi.org/10.4169/math.mag.90.1.48

[7] Palais, B., Palais, R., Rodi, S. (2009). A disorienting look at Euler's theorem on the axis of a rotation. *Amer. Math. Monthly.* 116(10): 892–909. doi.org/10.4169/000298909X477014
[8] Frey Jr., A., Singmaster, D. (1982). *Handbook of Cubik Math*. Hillside: Enslow Publishers.
[9] West, D. B. (1996). *Introduction to Graph Theory*. Upper Saddle River, NJ: Prentice-Hall, Inc.
[10] Wilson, R. J. (1996). *Introduction to Graph Theory*. Essex: Pearson Education Limited.

Summary. We enumerate all Hamilton circuits on the dual graph of the dodecahedron, group them according to symmetry, and then discuss which of them may be realized as snake patterns on the Megaminx puzzle. We find that there are eleven shapes that may be obtained without taking the puzzle apart (legal shapes) and ten more that can be achieved if you swap two pieces (illegal). With color taken into account, there are 700 ways to put a snake on the Megaminx, 280 of them legal.

THERESE AGLIALORO has now graduated with a mathematics degree from the University of Dallas, but was a freshman when making the discoveries described in this paper. She now works as a software developer for Epic Systems, but still enjoys thinking about math in her spare time.

ROBERT HOCHBERG (MR Author ID: 335758) received his Ph.D. from Rutgers, and is now an associate professor of mathematics and computer science at the University of Dallas. His interests include research in graph theory and combinatorics, and spending time with the most beautiful poetry.

Do the p-adic Integers Converge Somewhere?

ŞAHİN KOÇAK
Anadolu University
Eskişehir, Turkey
skocak@anadolu.edu.tr

YUNUS ÖZDEMİR
Eskişehir Technical University
Eskişehir, Turkey
yunuso@eskisehir.edu.tr

GÖKÇE ÖZKAYA
Eskişehir Technical University
Eskişehir, Turkey
gokceozkaya@eskisehir.edu.tr

More than a century after their invention by Kurt Hensel, the p-adic numbers (where p is a prime number) still have a mysterious aura surrounding them. Though they might seem counterintuitive and somewhat esoteric, they are actually simpler and more aesthetic, in some respects, than the reals. For example, consider the p-adic integers. We can conceive of them as a kind of fractal consisting of p clusters consisting of p clusters consisting ..., as in the poem of Jonathan Swift:

So, naturalists observe, a flea
Hath smaller fleas that on him prey
And these have smaller still to bite 'em
And so proceed ad infinitum

Let us enumerate the first-level clusters by $0, 1, \ldots, p-1$; the second-level clusters by $00, 01, \ldots, 0(p-1)$; $10, 11, \ldots, 1(p-1)$; etc., and the nth-level clusters by $a_0a_1 \ldots a_{n-1}$, where each $a_i \in \{0, 1, \ldots, p-1\}$ (see Figure 1). As this process proceeds ad infinitum, we can imagine that these shrinking clusters tend to a destiny $a_0a_1a_2 \ldots a_n \ldots$, a formal infinite word of letters taken from the alphabet $\{0, 1, \ldots, p-1\}$. These infinite words will be our p-adic integers.

We can identify a word $a_0a_1 \ldots a_n000 \ldots$, where eventually all letters become zero, with the ordinary integer $a_0 + a_1p + \cdots + a_np^n$. We view the other words as new formal objects and write them as

$$a_0 + a_1p + \cdots + a_np^n + \cdots .$$

So, we have a well-defined set, denoted by $\mathbb{Z}_p$, and we define a metric d on this set as follows: Let

$$a = a_0a_1a_2 \ldots \qquad \text{and} \qquad b = b_0b_1b_2 \ldots \in \mathbb{Z}_p.$$

Assume that $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$, ..., $a_{k-1} = b_{k-1}$, and $a_k \neq b_k$. Then we set $d(a, b) = \frac{1}{p^k}$. This gives a metric on $\mathbb{Z}_p$, which is a so-called ultrametric since it satisfies an inequality stronger than the triangle inequality:

$$d(a, b) \leq \max\{d(a, c), d(b, c)\}$$

Math. Mag. **95** (2022) 32–36. doi:10.1080/0025570X.2022.2001262
MSC: 51F99, 11E95

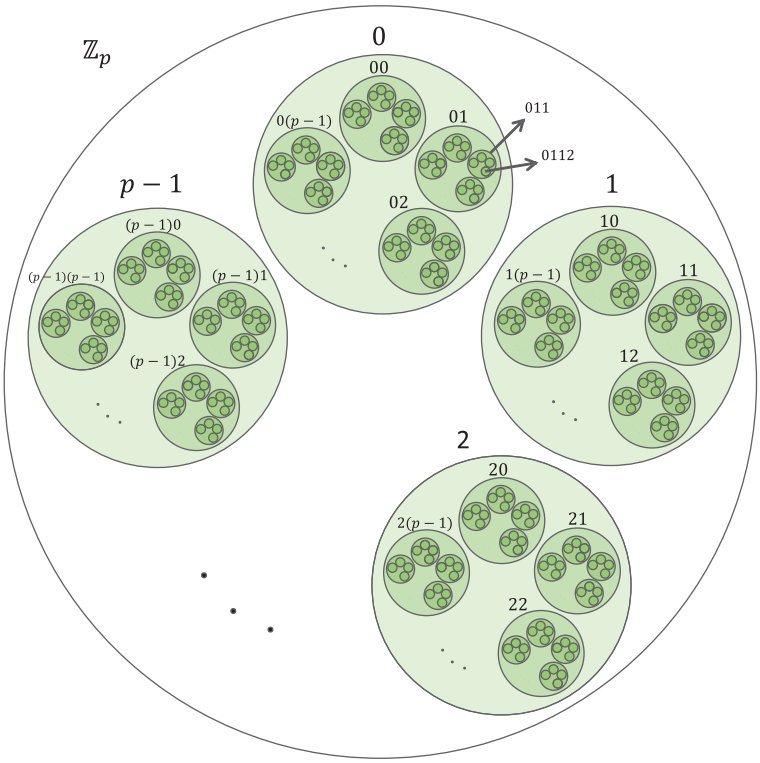


Figure 1 The fractal structure of p-adic integers.

There is a natural ring structure on the p-adic integers. Addition can be defined by term-wise addition with "carries." For example,

$$1000\ldots + (p-1)(p-1)(p-1)\ldots = 000\ldots,$$

so that $(p-1)(p-1)(p-1)\ldots$ is the additive inverse of 1, i.e. -1. We invite the reader to define the multiplication of p-adic integers. (Hint: Use a device like the Cauchy-product of series). We recommend the beautiful book by Svetlana Katok for an introduction to p-adic numbers [2].

Gromov–Hausdorff metric

Given a metric space, there is a distance between two points in it. Given two metric spaces, is there a distance between them? We mean of course, can one define a meaningful distance between them?

We will be concerned with compact metric spaces since the metric space $\mathbb{Z}_p$ of p-adic integers can be shown to be compact. Now, let X and Y be two compact metric spaces. Let us first assume that they are subspaces of a third metric space Z. There is certainly an $\varepsilon > 0$, such that the ε-neighborhood of X contains Y and the ε-neighborhood of Y contains X (see Figure 2). The infimum of such ε is called the Hausdorff distance of X and Y within Z. It was a clever stroke of Gromov to extend this definition to arbitrary pairs of metric spaces (see Burago, Burago, and Ivanov [1]), thereby making the definition more intrinsic, i.e., depending only on the given metric spaces.

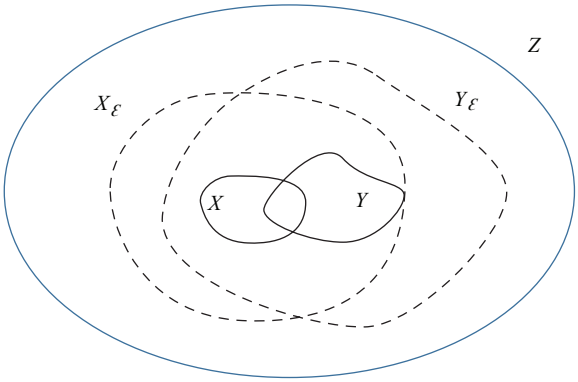


Figure 2 Hausdorff distance of X and Y within Z.

So, let X and Y be two compact metric spaces. Embed them isometrically into all possible metric spaces Z. (We leave it as an exercise to show there must be at least one such space Z). Compute the Hausdorff distance of the embedded copies of X and Y within Z. Then take the infimum over all possible Z. This gives a metric (denoted by d_{GH}) on the "space" of compact metric spaces. (One can eliminate set-theoretic difficulties by considering only "isometry classes" of compact metric spaces and taking a single representative from each class.)

Admittedly, it is not a simple matter to compute the Gromov-Hausdorff distance between two compact metric spaces, but there are easy tools for obtaining bounds on this distance. To recall one of them, we first define the notion of an ε-isometry between two metric spaces.

A subset S of a metric space (X, d) is called an ε-net, if

$$\operatorname{dist}(x, S) = \inf_{s \in S} d(x, s) \leq \varepsilon$$

for all $x \in X$.

The distortion of a map $f : X \to Y$ between two metric spaces (X, d_X) and (Y, d_Y) is defined by

$$\operatorname{dis} f = \sup_{x_1, x_2 \in X} |d_Y(f(x_1), f(x_2)) - d_X(x_1, x_2)|.$$

Now, a (not necessarily continuous) map $f : X \to Y$ is called an "ε-isometry" if $\operatorname{dis} f \leq \varepsilon$ and $f(X)$ is an ε-net in Y. The following property from the excellent book of D. Burago, Y. Burago and S. Ivanov [1] will be useful here.

Proposition 1. Let X and Y be (compact) metric spaces and $\varepsilon > 0$. Then,

(1) If $d_{GH}(X, Y) < \varepsilon$, then there exists a 2ε-isometry from X to Y.
(2) If there exists an ε-isometry from X to Y, then $d_{GH}(X, Y) < 2\varepsilon$.

The sequence of p-adic integers is not convergent

We want to show that the sequence of metric spaces of p-adic integers is not convergent in the metric space of compact metric spaces with the Gromov-Hausdorff metric. We do this by showing that this sequence is not a Cauchy sequence.

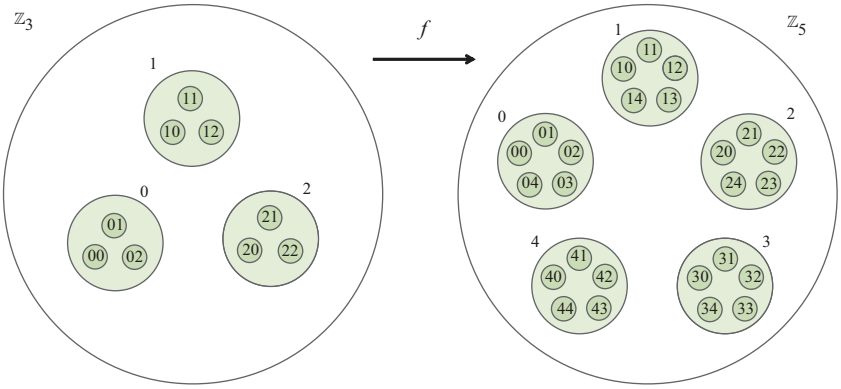


Figure 3 A map $f : \mathbb{Z}_3 \to \mathbb{Z}_5$.

Lemma 1. $d_{GH}(\mathbb{Z}_p, \mathbb{Z}_q) \geq \frac{1}{8}$ for any two prime numbers p, q $(p < q)$.

Proof. By the above proposition, it will be enough to show that there does not exist a $\frac{1}{4}$-isometry between $\mathbb{Z}_p$ and $\mathbb{Z}_q$. Consider any map $f : \mathbb{Z}_p \to \mathbb{Z}_q$. (For example, see Figure 3).

If we assume that the image of every first-level cluster in $\mathbb{Z}_p$ is contained in a first-level cluster of $\mathbb{Z}_q$, then there will be at least one first-level cluster of $\mathbb{Z}_q$, whose elements will not appear as image points under f. This will give $\operatorname{dist}(y, f(\mathbb{Z}_p)) = 1$ for such a point $y \in \mathbb{Z}_q$ and thus the map f can not be a $\frac{1}{4}$-isometry.

As a second possibility, if there exists a first-level cluster of $\mathbb{Z}_p$ which is not mapped wholly into a first-level cluster of $\mathbb{Z}_q$, then there will be at least two points x_1, x_2 of this cluster with $d_{\mathbb{Z}_q}(f(x_1), f(x_2)) = 1$. But $d_{\mathbb{Z}_p}(x_1, x_2) \leq \frac{1}{p}$ and hence

$$|d_{\mathbb{Z}_q}(f(x_1), f(x_2)) - d_{\mathbb{Z}_p}(x_1, x_2)| \geq \frac{1}{2}.$$

Thus, the distortion of f is at least $\frac{1}{2}$ and f cannot be a $\frac{1}{4}$-isometry. ■

Can we persuade the p-adic integers to converge?

Consider the sequence of balls $B_n = \{x \in \mathbb{R}^2 \mid \|x\| \leq n\}$ in the Euclidean plane. They do not converge with respect to the Gromov-Hausdorff metric. But everybody would wish them to converge somehow to the Euclidean plane itself. Likewise, it might be possible that the p-adic integers $\mathbb{Z}_p$ could converge to some non-compact metric space in some appropriate sense. It would be the second best thing, if a non-convergent sequence of compact metric spaces at least converges to a non-compact metric space.

To make the sequence of Euclidean balls converge to the Euclidean plane, Burago, Burago, and Ivanov [1] propose an extended notion of Gromov-Hausdorff convergence. However, to avoid the pointed metric spaces they use, we want to suggest an alternative concept of convergence:

Definition. Let $((X_n, d_n))_{n=1}^{\infty}$ be a sequence of compact metric spaces, and let (X, d) be a (non-compact) metric space. We say that this sequence converges to X, if there is a filtration

$$A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots \subset X$$

of X (with $\cup_{n=1}^{\infty} A_n = X$) such that, for any $\varepsilon > 0$ there exists an $n_0 \in \mathbb{N}$ with $d_{GH}(X_n, A_n) < \varepsilon$ for $n \geq n_0$.

We now show that the p-adic integers do converge in this sense!

Proposition 2. The sequence of compact metric spaces of p-adic integers $\mathbb{Z}_p$, ($p = 2, 3, 5, \ldots$) (i.e. the sequence $((X_n, d_n))_{n=1}^{\infty}$ with $X_n = \mathbb{Z}_{p_n}$, where p_n is the nth prime number and d_n is the p_n-adic metric) converges to the metric space of natural numbers $\mathbb{N}$ (including 0) with the discrete metric $d(m, n) = 1$ for $m \neq n$.

Proof. Consider the filtration $\{A_p\}$ of $\mathbb{N}$ with $A_p = \{0, 1, \ldots, p-1\}$. It will be enough to show that $d_{GH}(\mathbb{Z}_p, A_p) < 2/p$.

The map $f_p : \mathbb{Z}_p \to A_p$ that sends the ith first-level cluster of $\mathbb{Z}_p$ to the element $i \in A_p$ is a $\frac{1}{p}$-isometry:

(i) $f_p(\mathbb{Z}_p)$ is obviously a 0-net in A_p (since f_p is onto).

(ii) The distortion of f_p equals $\frac{1}{p}$. To see this, note that if $x_1, x_2 \in \mathbb{Z}_p$ belong to different first-level clusters, then

$$d_{\mathbb{Z}_p}(x_1, x_2) = 1 \qquad \text{and} \qquad d(f_p(x_1), f_p(x_2)) = 1.$$

If $x_1, x_2 \in \mathbb{Z}_p$ belong to the same first-level cluster, then $d_{\mathbb{Z}_p}(x_1, x_2) \leq \frac{1}{p}$ (this bound is attained) and $d(f_p(x_1), f_p(x_2)) = 0$.

Thus, $\operatorname{dis} f_p = \frac{1}{p}$ and consequently, by Proposition 1, $d_{GH}(\mathbb{Z}_p, A_p) < 2/p$. ■

Given this proposition, one is tempted to speculate that the p-adic metric with respect to the "infinite prime" on $\mathbb{N}$ (or $\mathbb{Z}$) should be $d(m, n) = 1$ for $m \neq n$, instead of the traditional $d(m, n) = |m - n|$. This is all the more natural as $d_p(m, n) = 1$ for a prime p if and only if m and n are coprime, and any two integers are "coprime" with respect to the "infinite prime"!

REFERENCES

[1] Burago, D., Burago, Y., Ivanov, S. (2001). *A Course in Metric Geometry*. Providence, RI: Amer. Math. Soc.
[2] Katok, S. (2007). *p-adic Analysis Compared with Real*. Providence, RI: Amer. Math. Soc.

Summary. We show that the sequence of metric spaces of p-adic integers $\mathbb{Z}_p$ converge with respect to a modified version of Gromov–Hausdorff convergence to a countably-infinite discrete metric space.

ŞAHİN KOÇAK (MR Author ID: 244537) received his Ph.D. in mathematics from the Heidelberg University in 1979. He is fond of fractals, and after some forty years at the Anadolu University, he founded a Mathematics Village in the mountains near Eskişehir, Turkey. (www.cakilarasimatematikkoyu.com)

YUNUS ÖZDEMİR (MR Author ID: 822834) received his Ph.D. in mathematics from the Anadolu University, Turkey, in 2008. He is currently a professor at the Eskişehir Technical University, Turkey. His main mathematical interests include fractal and metric geometry. He enjoys spending time with his family, driving a car and playing football.

GÖKÇE ÖZKAYA received her masters degree in mathematics from the Eskişehir Technical University, Turkey, in 2019. Her main research interests are in geometry and topology. She loves to be in black and enjoys drinking filtered coffee anytime of the day.

A Search for Champion Boxers

TIEN CHIH
Montana State University-Billings
Billings, MT 59101
tien.chih@msubillings.edu

DEMITRI PLESSAS
Independent Researcher
Kenmore, WA 98028
demitri.plessas@gmail.com

> Once the fans of history get an idea of the people he beat, then they will get a better perspective of him, I'm sure. He's got an All-Star list of victories. So they're gonna think '. . . damn he beat all these guys?' People are gonna know. He should be in the top 5, top 3 and stuff. –Mike Tyson on Evander Holyfield.
>
> (From the documentary *Chasing Tyson* [3])

The 2015 television documentary *Chasing Tyson* [3] describes the career of four-time world heavyweight champion Evander Holyfield. One of the film's central themes was how, due to his low-key style and humble attitude, the boxing public did not accept Holyfield as a top fighter deserving of the accolade "world heavyweight champion." However, former heavyweight champion and widely accepted all-time great Mike Tyson is quoted above at the end of the film, making the case for Holyfield's greatness. Although not particularly flashy in or outside the ring, Holyfield's achievement in boxing may be measured in an objective way: by the list of quality boxers he defeated.

This argument is reminiscent of Google's PageRank algorithm [2, 11]. PageRank is an algorithm that assigns numerical values to web pages, based on the pages that link to it and *their* value. Receiving an incoming link from a high-value website carries more value than a link from a low-level one. The analogy to boxing is clear: A boxer's status is determined by the opponents they beat, with victories over higher-quality opponents holding more value than victories over weaker opponents. Of course, the values of those other boxers are determined the same way, suggesting that a global analysis of all fighters is necessary.

We describe the mathematics behind PageRank and give a scheme for adapting it to boxing. We illustrate the algorithm's behavior in several scenarios. Then 20 top heavyweight boxers from the 1990s are selected to see if Mike Tyson was justified in his praise of Evander Holyfield. We then make a comparison to the Colley Method for ranking sports teams, which also depends on the quality of opponents, and we describe some possible extensions.

Shifting values, stochastic matrices, and Markov chains

Our model will use Markov chains that are row-stochastic and act on the right of row vectors. In order to model our boxing network and apply the PageRank algorithm, it is necessary to keep in mind two different views of what a Markov chain can represent.

Math. Mag. **95** (2022) 37–48. doi:10.1080/0025570X.2022.2000819

As an example, consider the Markov chain

$$M = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

which has a weighted digraph representation (Figure 1). One way of interpreting this matrix is as a value shifting scheme. Suppose that Alice, Bob and Chen are three people, each with some amount of water. They pass water to each other in such a way that at each phase of the process, Alice gives $1/3$ of her water away to Bob and Chen, Bob gives away $1/4$ of his water to Alice and Chen, and Chen gives half his water to Alice.

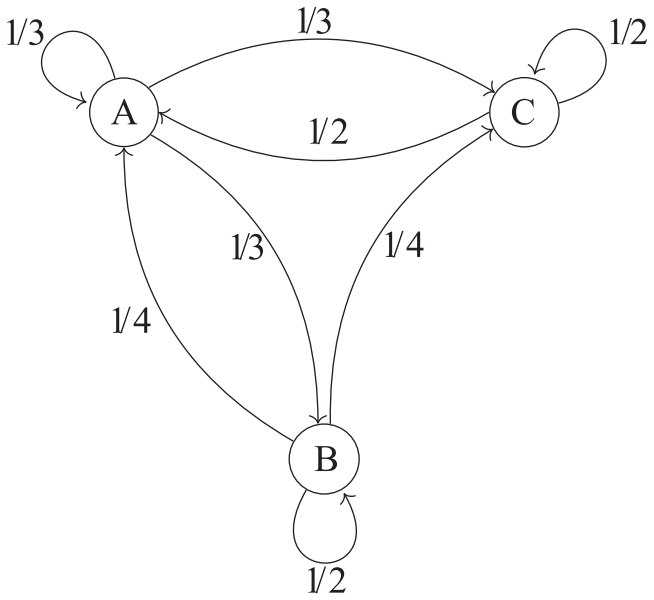


Figure 1 Digraph representation of a Markov chain.

It would then be reasonable to ask: Is there some quantity of water per person such that after each phase of the process, each person's quantity remains unchanged? In other words, are there solutions to

$$\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \end{bmatrix} = \begin{bmatrix} a & b & c \end{bmatrix},$$

or, equivalently, is there a left eigenvector for eigenvalue $\lambda = 1$? We can see by inspection that

$$\begin{bmatrix} 3/8 & 1/4 & 3/8 \end{bmatrix}$$

is one such solution.

This eigenvector can also be used to answer the question: If Alice were to start with all the water, which we can take to be one cup, and if this process were repeated

indefinitely, then would each person's quantity of water converge to some value? This quantity would be represented as

$$\lim_{n\to\infty} [1, 0, 0]M^n.$$

We can see that as we iterate this process

$$\begin{bmatrix}1 & 0 & 0\end{bmatrix}\begin{bmatrix}1/3 & 1/3 & 1/3\\ 1/4 & 1/2 & 1/4\\ 1/2 & 0 & 1/2\end{bmatrix} = \begin{bmatrix}1/3 & 1/3 & 1/3\end{bmatrix},$$

$$\begin{bmatrix}1 & 0 & 0\end{bmatrix}\begin{bmatrix}1/3 & 1/3 & 1/3\\ 1/4 & 1/2 & 1/4\\ 1/2 & 0 & 1/2\end{bmatrix}^2 = \begin{bmatrix}13/36 & 5/18 & 13/36\end{bmatrix}, \quad \text{and}$$

$$\begin{bmatrix}1 & 0 & 0\end{bmatrix}\begin{bmatrix}1/3 & 1/3 & 1/3\\ 1/4 & 1/2 & 1/4\\ 1/2 & 0 & 1/2\end{bmatrix}^{100} \approx \begin{bmatrix}3/8 & 1/4 & 3/8\end{bmatrix}.$$

We achieve the same values we had before.

Our second way of interpreting this matrix is as a representation of stochastic probability. An ant standing on node A has a $1/3$ chance of staying on A and a $1/3$ chance of moving to nodes B or C. An ant on node B has a $1/2$ chance of staying on node B and a $1/4$ chance of moving to either A or C. If the ant is on node C, then it has a $1/2$ chance of staying on node C and a $1/2$ chance of moving to A. It is reasonable to ask a question like: If an ant starts on a particular node, say A, then as this process iterates, what is the ant's probability of landing on each node? Since the probability of the ant landing on any given node after n steps is

$$\begin{bmatrix}1 & 0 & 0\end{bmatrix}\begin{bmatrix}1/3 & 1/3 & 1/3\\ 1/4 & 1/2 & 1/4\\ 1/2 & 0 & 1/2\end{bmatrix}^n,$$

we can surmise that the probability that the ant lands on any particular node, as n tends to infinity, is $3/8$ for A, $1/4$ for B and $3/8$ for C.

So the key to our ranking system is noting that these Markov matrices represent both stochastic probabilities and value shifting schemes. Moreover, reiteration of either process results in converging to a *stable state*, or left eigenvector for eigenvalue $\lambda = 1$.

Modeling a boxing network

We can see why we would want to use stochastic matrices to model the ranking of boxers. After all, our entire premise is that when a boxer defeats an opponent, the value of their opponent is transferred to them, and the more valuable the opponent, the more it should enhance the status of the victor.

Let us consider a simple scenario: Suppose there are three boxers creatively named A, B, and C. First suppose that A and B fight, and B emerges victorious. They then each go on to fight C and lose. We can see a visualization of who loses to whom in Figure 2.

Now, if there is any justice in the universe, a reasonable scheme to rank these boxers should rank C ahead of B ahead of A. Let us think of this as a value shifting scheme,

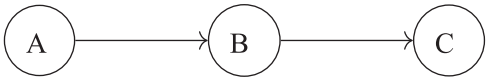


Figure 2 A simple boxing network.

where each boxer is carrying around a glass of water, and losing to other boxers forces them to share that water. Since A loses to B, A has to give B their water, and similarly, B loses to C. If we represent this as a matrix, we obtain

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ ? & ? & ? \end{bmatrix}.$$

It is not entirely clear what is to be done about boxer C, since, after all, boxer C does not lose to anyone. Should we allow boxer C to keep all their water? If we do, then this is equivalent to saying that boxer C defeated himself. Does this really make sense?

It is useful to imagine this matrix representing the probabilities in a Markov chain. We know A lost to B, and barring any further data, we can conclude that A has a 100% chance of losing to B. Similarly, B has lost to C 100% of the times they fought. What about C, though? Since C has suffered no losses, technically everyone has an equal chance to defeat C, including himself! After all, A, B **and** C all have defeated C an equal number of times. With this in mind, we can fill out this last row:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

How do we rank our boxers then? Well, just as with our water sharing friends Abby, Ben and Chen, we have a value-sharing scheme represented by a stochastic matrix. This matrix has an eigenvector with eigenvalue 1, representing a stable state for the amount of water each person approaches after iterating this process. Since we now have a water sharing scheme for our three boxers, we can solve for this state in a now familiar way:

$$\begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} = \begin{bmatrix} a & b & c \end{bmatrix},$$

and therefore,

$$\frac{1}{3}c = a, \qquad a + \frac{1}{3}c = b, \qquad \text{and} \qquad b + \frac{1}{3}c = c.$$

which has solution $a[1, 2, 3]$. Note that for any positive value we choose for a, we always have that b will have greater value, and c greater value still. Thus, we obtain a sensible ranking of these boxers.

For a more in-depth example, suppose there are five boxers, A, B, C, D, E. Boxer A loses to B, who then loses to C. Then E fights B to a draw, and D fights A and wins. Afterwards, E and B have a rematch and E wins. This is all represented in Figure 3.

How would we rank these boxers? Would D be ranked the same as C as they are both undefeated? How does E compare to everyone?

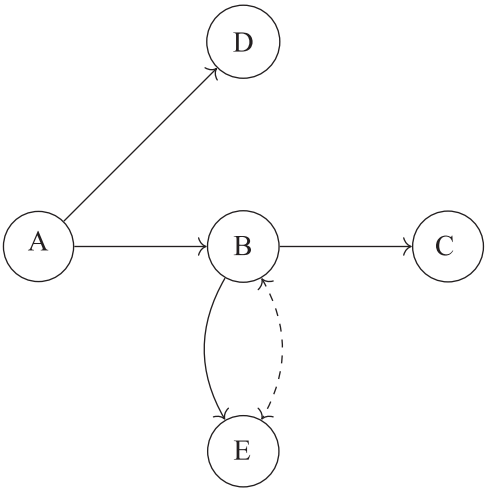


Figure 3 A more complex boxing network.

Here we address some of this increased complexity. To begin with, a draw should not be ignored by our scheme. Fighting a champion to a draw implies good things about a fighter. In our scheme, we will count each draw as half a loss for each fighter. Doing so gives us the following raw data matrix:

$$R = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 3/2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \end{bmatrix}.$$

This is certainly *not* a stochastic matrix. It is simply a record of who lost to whom, and, in a sense, how many times. We also have 0's in lieu of the question marks ? from earlier in this section since this is not meant to represent a stochastic matrix, and since C, D are undefeated. Here we note that since B tied E once and lost to E once, the B, E entry is $1/2 + 1 = 3/2$, and similarly the E, B entry is $1/2$.

Our first step is to normalize the results of multiple fights. Rematches are a common occurrence in boxing, especially if the public or one of the parties disputes the results of a previous match. Ultimately, our goal is to compare the boxers to each other, and two boxers who fight against each other multiple times should not have their mutual comparison have an inordinate impact on their ranking. In our case, boxers B and E fought twice, with B achieving 1/2 a victory, and E achieving 3/2's victories. To normalize this, we take the total number of fights and divide the number of victories by that total. Thus, B would be understood to have $(1/2)/2 = 1/4$ victories over E, and E would have $(3/2)/2 = 3/4$ victories over B. Our normalized matrix is this:

$$N = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 3/4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 & 0 \end{bmatrix}.$$

The next step is to stochasticize this matrix. For a boxer with ties and losses, we simply divide their row by their row sum, guaranteeing a row sum of 1. Boxers with

only victories would have 0 rows, and since no one has actually defeated them, everyone has an equal chance of defeating them, making each entry of their row 1/5:

$$S = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 4/7 & 0 & 3/7 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

This matrix makes intuitive sense. Since A loses to C and D they split their value. As C and E clearly defeated B, but E also suffered a draw, C is able to take more value from B than E is. Finally, E only suffers a draw to B, but if E were to lose to anyone, it would be B. When we find the stable state, we obtain

$$t\begin{bmatrix} 1 & 4.375 & 3.5 & 1.5 & 2.875 \end{bmatrix}$$

where t is a real number. To make sense of this ranking, it is clear that A would be ranked the lowest. Although C and D are both undefeated, C has beaten a boxer with a better performance than the one D defeated, so C has a higher ranking than D. Although both C and E bested B, they have only fought 1fighter, while B has a win and a partial victory against 2 fighters, allowing more value to flow to B. If E or C wishes to increase their ranking, they would need to fight boxers other than B.

A possible complication

Suppose we once again have boxers A, B, C, D, E, where A loses to B who loses to C who loses to A. Moreover, D and E fight and draw. How should we rank these boxers?

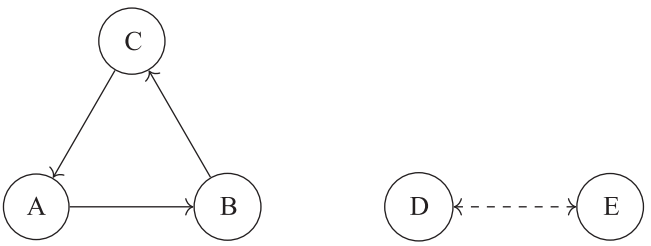


Figure 4 A disconnected boxing network.

It is clear that no boxer seems particularly dominant. In fact, it would not be a stretch to say that all the boxers should be ranked equally. When we compute the raw data, normalized, and stochastic matrices, we obtain:

$$R, N = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \qquad S = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

However, when we try to find our stable states, we notice the vector

$$\begin{bmatrix} 3 & 3 & 3 & 1 & 1 \end{bmatrix}$$

is a stable state for this system. But

$$\begin{bmatrix} 1 & 1 & 1 & 100 & 100 \end{bmatrix}$$

is also a stable state. What is going on?

Considering Figure 4, we can see that there are basically two separated networks here. Each network could be dealt with on its own, and within those networks, it is easy to see everyone has equal value. However, as there are no fights between fighters A, B, C and D, E, there is no way to compare the value of those two groups. The fighters in the first group could be much better or much worse than the fighters in the second, but without a match or comparison, there is no way to tell. This is reflected in our matrix S. We see that S is a block matrix for which

$$\begin{bmatrix} s & s & s & t & t \end{bmatrix}$$

will be an eigenvector with eigenvalue 1 for any real numbers s and t

$$S = \left[\begin{array}{ccc|cc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right].$$

However, in this situation, it is still necessary to be able to produce a ranking.

Our solution is to introduce J, the all 1's matrix. It is clear that if J is an $n \times n$ matrix, then $\frac{1}{n}J$ is a stochastic matrix. Then given any p, we have $0 \leq p \leq 1$, and we observe that

$$M = (1 - p)S + p(\frac{1}{n}J)$$

is stochastic as well. Moreover, if p is positive, then every entry of M will also be positive. This value p is called the *damping factor* [2].

Why do we create this matrix M? If p is sufficiently small, then $\frac{p}{n}J$ is unlikely to change the relative rankings that would have been established by S. More importantly, this allows us to invoke the following theorem, which only applies to matrices with positive, rather than non-negative, values.

Theorem 1 (Perron-Frobenius Theorem [9])**.** *If M is a row stochastic matrix where each entry is positive then:*

1. 1 *is an eigenvalue of multiplicity one.*
2. 1 *is the largest eigenvalue. All the other eigenvalues have an absolute value smaller than 1.*
3. *The eigenvectors corresponding to the eigenvalue* 1 *have either only positive entries or only negative entries. In particular, for the eigenvalue* 1*, there exists a unique eigenvector with the sum of its entries equal to* 1.

Items 1 and 3 are of particular importance since they guarantee the existence of a unique ranking. A value of $p = 0.15$ is typical for many implementations of this program. In our case, if $p = 0.15$, then

$$M = (1 - 0.15)S + (0.15)\frac{1}{5}J = \begin{bmatrix} 0.03 & 0.88 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.88 & 0.03 & 0.03 \\ 0.88 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.88 \\ 0.03 & 0.03 & 0.03 & 0.88 & 0.03 \end{bmatrix},$$

which has stable-state $t\begin{bmatrix}1 & 1 & 1 & 1 & 1,\end{bmatrix}$ where t is a real number, as we suspected.

The final algorithm

Putting everything together, our final algorithm has the following steps:

1. Start with a raw data matrix R, where $R_{i,j}$ is the number of times boxer i losses to boxer j, with draws counting as a half-loss.
2. Create a normalized data matrix N, where $N_{i,j} = R_{i,j}/(R_{i,j} + R_{j,i})$ if $R_{i,j} + R_{j,i} \neq 0$, and $N_{i,j} = 0$ otherwise.
3. Create a stochastic data matrix S, where $S_{i,j} = N_{i,j}/(\sum_{k=1}^{n} N_{i,k})$ if $\sum_{k=1}^{n} N_{i,k} \neq 0$ and $N_{i,j} = 1/n$ otherwise.
4. Choose a damping factor p, $0 < p < 1$ and create $M = (1 - p)S + p\frac{1}{n}J$.
5. Find a positive left eigenvector of M that has eigenvalue 1.

This roughly describes the implementation of PageRank [2, 11]. When used to rank web pages, web pages values are determined by incoming links and the respective value of those pages. In the boxing context, some modification was required to address the issues of multiple fights and draws, which are not relevant to ranking websites. More on using PageRank in sports can be found in Govan, Meyer, and Albright [10], and Langville and Meyer [11, 12].

Was Mike Tyson right or not?

Evander Holyfield's best-remembered fights took place in the 1990s, and so looking at a list of the top 20 boxers of the 1990 (as assessed by the website "heavyweightaction.com" in 2016, though this website is no longer active), we find

1. Lennox Lewis
2. Riddick Bowe
3. Evander Holyfield
4. Mike Tyson
5. George Foreman
6. Ike Ibeabuchi
7. Michael Moorer
8. Buster Douglas
9. Oliver McCall
10. Larry Holmes
11. Ray Mercer
12. David Tua
13. Tommy Morrison
14. Frank Bruno
15. Razor Ruddock
16. Michael Grant
17. Andrew Golata
18. Tony Tucker
19. Tim Witherspoon
20. Henry Akinwande

The website gives its reasons for this particular ranking. However, we will determine our own ranking with our algorithm. Looking at fights that took place between these fighters in the 1990 [14], we obtain the following network of fights (Figure 5):

Then, allowing $p = 0.15$, the stable state for $M = (1 - p)S + p\frac{1}{20}J$ is approximately:

$$t[1 \quad 0.394 \quad 1.104 \quad 0.188 \quad 0.215 \quad 0.089 \quad 0.293 \quad 0.128 \quad 0.726 \quad 0.086$$
$$0.134 \quad 0.048 \quad 0.153 \quad 0.295 \quad 0.048 \quad 0.068 \quad 0.069 \quad 0.295 \quad 0.048 \quad 0.174]$$

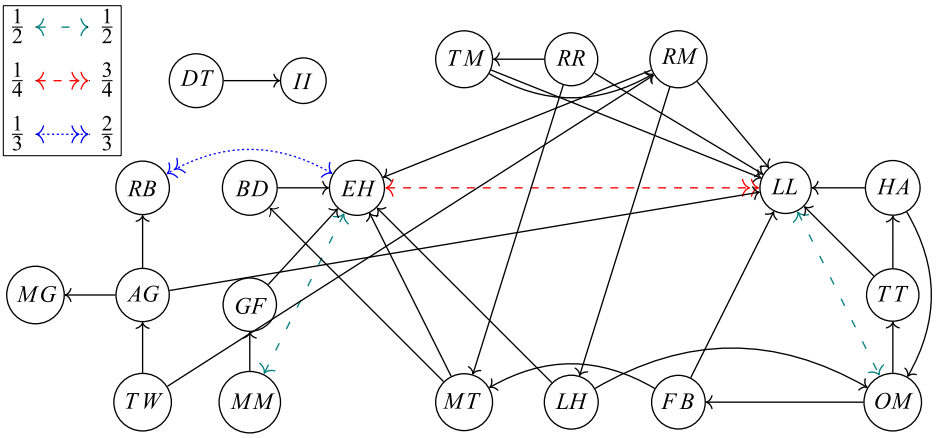


Figure 5 The boxing network for the top 20 boxers of the 1990.

It follows that by our method, the ranking is

1. Evander Holyfield
2. Lennox Lewis
3. Oliver McCall
4. Riddick Bowe
5. Frank Bruno
5. Tony Tucker
7. Michael Moorer
8. George Foreman
9. Mike Tyson
10. Henry Akinwande
11. Tommy Morrison
12. Ray Mercer
13. Buster Douglas
14. Ike Ibeabuchi
15. Larry Holmes
16. Andrew Golata
17. Michael Grant
18. Razor Ruddock
18. David Tua
18. Tim Witherspoon

As Mike Tyson intimated, Holyfield has an impressive list of victories, including over Mike Tyson himself. Moreover, no fighter who bested Holyfield was able to do so unilaterally. The only fighters who beat Holyfield were Riddick Bowe, Michael Moorer, and Lennox Lewis. Holyfield lost to Bowe in 1992, defeated Bowe in 1993, and lost to him a second time in 1995. He lost to Moorer in 1997 and avenged his loss in 1997. His initial fight against Lennox was a draw in 1999, but he was defeated by Lennox later that year [**14**]. So even though these fighters defeated Holyfield, the draws and comeback wins make those only partial victories and allows for some of their value to flow back to Holyfield. Thus, by our analysis, Mike Tyson was right, and Holyfield's preeminence should not be ignored.

Comparison to the Colley method

PageRank is not the only matrix-based ranking algorithm to be applied to sports. An alternative approach is the Colley method developed, by Astrophysicist Dr. Wessley Colley [**5**, **6**, **12**]. The general principles of this method are as follows:

1. The algorithm is *bias-free*, meaning it only uses win/loss/tie information to determine the rankings. It does not consider ad hoc adjustments likes leagues, points scored, or rounds won.
2. It follows a principle of *conservation* in that each team/athlete starts with a value of 1/2. Each time a team/athlete defeats another, part of the loser's value is transferred

to the winner. The stronger the opponent defeated, the more value is transferred to the winner.

As we can see, this method shares a lot of the same principles as PageRank. The ranking vector is found by solving for $\mathbf{r}$, for $C\mathbf{r} = \mathbf{b}$ where

$$C_{i,j} = \begin{cases} 2 + t_i & i = j \\ -n_{i,j} & i \neq j \end{cases}, \qquad b_i = 1 + \frac{w_i - l_i}{2}.$$

Here, t_i represents the total number of games by team/athlete i, w_i, l_i are their wins and losses, and $n_{i,j}$ represents the number of matches between team/athletes i and j.

Perhaps, the most significant difference between the Colley method and PageRank is that while Colley transfers some value from the loser to the winner in each defeat, PageRank uses the win/loss/tie information to establish a hierarchy between the team/athletes. PageRank does not count the number of matches between two teams/athletes, only the normalized outcomes.

For example, suppose two unranked boxers, A and B, enter a match, and A defeats B. Consider the Colley method to find a ranking vector for these boxers. Prior to this match, we would have valued each boxer at $1/2$. After the match we would have

$$\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} r_A \\ r_B \end{bmatrix} = \begin{bmatrix} 3/2 \\ 1/2 \end{bmatrix}, \qquad \begin{bmatrix} r_A \\ r_B \end{bmatrix} = \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix}$$

A, by winning the match, has taken some of B's value for themselves. If they had a rematch, and if A won again, then we would obtain a new ranking vector:

$$\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} r_A \\ r_B \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} r_A \\ r_B \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$$

In this case, we would have even more of B's value transferred to A.

On the other hand, PageRank would not distinguish between the first and second wins. The first match establishes A as the dominant fighter over B. The rematch then merely confirms this dominance. In PageRank's eyes, it does nothing to change the relationship between A and B. In either case, we have:

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix}, \qquad \begin{bmatrix} a & b \end{bmatrix} = t \begin{bmatrix} 1 & 1/2 \end{bmatrix},$$

for t a real number. If we choose $t = 2/3$, then we obtain the same vector as the Colley method after back to back wins for A.

This leads to a second difference between these methods. As noted by Chartier, Kreutzer, Langville, and Pedings [4], the PageRank algorithm is much more sensitive to upset victories. This sensitivity is also related to the fact that PageRank does not distinguish between 1 or 5 or 100 victories between two opponents, whereas Colley transfers value from winners to losers in a more conservative manner. Moreover, since the damping factor chosen for the PageRank algorithm is somewhat arbitrary, we should note that the ranking vector produced by this process is for purposes of creating an ordinal ranking, and that the exact values of the vector are of little significance.

The ranking of the 90 Heavyweights by the Colley method is

1. Lennox Lewis
2. Riddick Bowe
3. Evander Holyfield
4. Oliver McCall
5. Buster Douglas
6. Tommy Morrison
7. Michael Grant
8. Ray Mercer
9. Ike Ibeabuchi

10. Mike Tyson
11. Frank Bruno
12. Henry Akinwande
13. Tony Tucker
13. Larry Holmes
15. Michael Moorer
16. George Foreman
17. Andrew Golata
18. David Tua
19. Razor Ruddock
20. Tim Witherspoon

Further questions

Although our ranking system showed that Evander Holyfield was ranked higher than Mike Tyson—and in fact Holyfield beat Tyson twice—very few people would say that Holyfield was a better fighter than Tyson. Many would argue that Holyfield fought Tyson at the tail end of Tyson's career, when Tyson was showing signs of age [**3**]. Certainly during the 1980's, when Tyson was considered to be at his best, every fight resulted in victory, which cannot be said about Holyfield in the 1990's.

Problem 1. Can one find a way to weight the value or importance of fights given where each fighter is in their career?

Problem 2. Is there a way to fairly compare boxers who may not have fought in the same time frame, or in overlapping time frames?

Our analysis here is close to binary, in that we have wins and losses with ties splitting the difference. However, commentators and boxing enthusiasts not only place value on who wins, but also on how they won. One of Mike Tyson's most infamous victories was against Michael Spinks in 1988, in which he knocked out the previously undefeated champion in 91 seconds [**8**].

Problem 3. Is there a way to weight the value of victories, knockouts versus decisions, factoring in point-spreads and round on knockouts (if they occur), to have a less binary analysis? How would this change to our algorithm compare to the Massey method of ranking? How would winning because your opponent bit off a part of your ear factor into this?

Acknowledgments The authors would like to thank the Dr. Michael Jones and the referees for their suggestions, which improved the paper.

Tien Chih would like to thank his son Maxton Dun Chih, born April 19th, 2016. Without the many sleepless nights following his birth, the author would not have watched the appropriate documentaries on both boxing [**3**] and mathematics [**7**] which inspired this paper.

This paper is dedicated to the memory of Muhammad Ali (January 17, 1942 - June 3, 2016). He was a champion of not only boxing but of justice and human rights as well.

REFERENCES

[1] Bondy, A., Murphy, U. S. R. (2008). *Graph Theory*. New York: Springer.
[2] Bryan, K., Leise, T. (2006). The $25,000,000,000 eigenvector. The linear algebra behind Google. *SIAM Rev*. 48(3): 569–581. doi.org/10.1137/050623280
[3] *Chasing Tyson*, documentary. (2015). Directed by Steven Cantor. ESPN 30-for-30.
[4] Chartier, T. P., Kreutzer, E., Langville, A. N., Pedings, K. E. (2011). *Sensitivity and Stability of Ranking Vectors*. *SIAM J. Sci. Comp.* 33(3): 1077–1102. doi.org/10.1137/090772745
[5] Chartier, T. P., Kreutzer, E., Langville, A. N., Pedings, K. E. (2010). Bracketology: How can math help? In: *Mathematics and Sports*, J. A. Gallian, ed. Providence, RI: Math. Ass. Amer. https://doi.org/10.5948/UPO9781614442004.006. October 2021.
[6] Colley, W. N. (2002). Colley's bias free college football ranking method: The Colley matrix explained. https://www.colleyrankings.com/matrate.pdf. October 2021.
[7] Du Sautoy, M. (2015). *The Secret Rules of Modern Living: Algorithms*. BBC Four.
[8] Gildea, W. (1988). Tyson-Spinks is more than a fight, it's history. *Los Angeles Times*, June 19, 1988.

[9] Godsil, C., Royle, G. (2001). *Algebraic Graph Theory*. New York: Springer.
[10] Govan, A. Y., Meyer, C. D., Albright, R. (2008). Generalizing Google's PageRank to rank National Football League teams. *SAS Global Forum 2008: Data Mining and Predictive Modeling*. https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/151-2008.pdf. October 2021.
[11] Langville, A. N., Meyer, C. D. (2006), *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton, NJ: Princeton University Press.
[12] Langville, A. N., Meyer, C. D. (2012). *Who's # 1?: The Science of Rating and Ranking*. Princeton, NJ: Princeton University Press.
[13] Lay, D. C., Lay, S. R., McDonald, J. J. (2016). *Linear Algebra and It's Applciations,* 5th ed. Boston, MA: Pearson.
[14] *The Boxing Database*. (2016). http://boxrec.com/. October 2021.

Summary. In the world of boxing, fighters are judged by the quality of their opponents. This way of determining value is reminiscent of Google's PageRank algorithm, which assigns values to websites based on the value of sites linked to it. We describe the mathematics behind PageRank, and then offer a scheme for adapting it to boxing. We illustrate the behavior of this algorithm in several scenarios. We then select 20 top heavyweight boxers from the 1990's and rank them using our method. A comparison is made to the Colley method of ranking. We conclude by describing some possible extensions of this method.

TIEN CHIH (MR Author ID: 1223485) earned a B.A. from the University of Hawaii at Hilo, and an M.A. and Ph.D. from the University of Montana. He enjoys his current appointment as Assistant Professor at Montana State University-Billings, where he works on integrating high impact practices into his courses, mentoring undergraduates, running a Math Circle, and conducting research on Categorical Graph Theory. He enjoys cooking for his wife and two sons.

DEMITRI PLESSAS (MR Author ID: 1223509) earned a B.S. from Montana Tech, and an M.A. and Ph.D. from the University of Montana. He loves blending mathematics and statistics to tackle unique modeling challenges as a Lead Data Scientist at Zulily. As the proud father of a son and a daughter, his greatest joy in life is seeing the world through their eyes.

The Diversity Discrepancy, Or Why Your Classmates Are More Like You

BRETT HEMENWAY
University of Pennsylvania
Philadelphia, PA, 19104
fbrett@cis.upenn.edu

DAVID HEMENWAY
Harvard University
Cambridge, MA 02138
hemenway@hsph.harvard.edu

Suppose a school has 100 students divided into two classes. One class has 99 students, and one class has only a single student (each student is in exactly one class). From the point of view of the administration, the expected class has 50 students (100 students, divided into two classes). From the point of view of the students, however, the expected class size is $98.02 = (.99 \cdot 99) + (.01 \cdot 1)$ (since 99% of the students are in a class of size 99, while 1% of the students are in a class of size 1). Both calculations are mathematically valid, and different circumstances call for different notions of the expected class size. But it may seem a little disingenuous from the student's perspective when the school's brochure advertises 50 students per class, while the average student experiences a class size of over 98.

Notice that these two notions of average will coincide exactly when the classes are of equal size. This is because the students are essentially taking a weighted average of the class size, where each class is weighted by its size, while the administration is weighting each class equally.

This phenomenon has been documented and analyzed in a variety of settings. For example, if you ask a random diner to count the number of other patrons in the restaurant they dined in last night, they should give you a higher estimate (on average) than if you ask a random restaurant owner. It can also explain why the movies you attend are often crowded, even though overall theater attendance is low, or why the people you see at the gym on a random visit are fitter than the average gym member [**6**]. On the cosmological scale, it can even explain why our galaxy is likely to be larger than average [**5**].

This is also the basis of the "friendship paradox," which states that on average your friends have more friends than you do [**4**]—this phenomenon has been verified using online social networks, and if you define a "friend" as someone you follow on Twitter, then on average your followers have more friends than you do, and this holds for more than 98% of Twitter users [**7**].

In this note, we argue that the same phenomenon applies to measures of diversity. Now, diversity can be defined in many ways. Here we use a simple definition, the percent (or absolute number) of the minority when there are only two types of people (e.g., Americans and Canadians, men and women, adults and children, Whites and Blacks). We assume that a group with a 20% minority is more diverse than one with a 10% minority. In this note, we show that most people attend classrooms or live in neighborhoods that are less diverse than one would expect—the majority perceives less diversity while the minority perceives more diversity. This is a mathematical fact, and it holds regardless of whether people are attracted to others with similar characteristics.

Math. Mag. **95** (2022) 49–58. doi:10.1080/0025570X.2022.2000263
MSC: 62D99
Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/0025570X.2022.2000263

Returning to the simple example above, suppose a school has 100 male students and 100 female students divided into two classes. Suppose one class has 99 boys and one girl, whereas the other class has 99 girls and one boy. If we ignored the boys altogether, then we would be in the situation described above (the average class has 50 girls, whereas the average girl is in a class with 98.02 girls).

The diversity discrepancy comes from conflating three different types of averaging (or different sampling procedures). Consider three natural measures of diversity, which we will label aggregate, overall and perceived.

(A) *Aggregate diversity:* Choose a random class uniformly from the set of classes, then examine the gender breakdown in that class.

(B) *Overall diversity:* Choose a random student uniformly from the population, and examine the gender breakdown in that student's class.

(C) *Perceived diversity:* Choose a random female (respectively male) uniformly from the population and examine the gender breakdown in her (respectively his) class.

The observation is simply that these three different measures of diversity can give strikingly different results. For example, perceived diversity (determined by choosing a random female) is most likely to find an average class that has the highest percentage of females, higher than either the aggregate diversity (choosing a random class) or overall diversity (choosing a random student). This is because when we choose a random female she is more likely to be a student from a class with a lot of females.

Sampling procedure (B) mimics the experience of a random individual, sampling procedure (C) mimics the experience of an individual in the population of interest (e.g. females); the results of sampling procedure (A) and (B) are what is usually reported (e.g., [**9**]), analyzed and used for decision-making. The "overall diversity," (B) as outlined above, coincides with the overall population diversity (ignoring classes altogether). Thus, if a school's population is ten percent Canadian, a randomly chosen student will (on average) be in a class that is ten percent Canadian.

The take-home message of this note is that, on average, individuals will experience classrooms, teams and neighborhoods that are more similar to them than one might expect from looking at aggregate statistics.

Why your neighborhood is less diverse than average

The diversity discrepancy explains why individuals are more likely to live in a county whose population is "more like them" than one might expect from the population-level statistics. There are economic, social and cultural factors that might cause people choose to live near their racial or ethnic peers, but this type of social clustering is independent of the point we are making.

Even if every person were randomly assigned to live in a certain county, we would still find that, on average, each person would find themselves living in a county whose population was more like them than might be expected based on looking at the population-level statistics. In the supplemental materials, we examine (through simulations) the diversity discrepancy when each citizen is randomly assigned to a county.

The fact that social and cultural factors might cause people to live with their peers will serve to increase this discrepancy, but it is not the underlying cause.

The following data (see Table 1) are from the 2010 US Census [**1**].

First, we notice that for most people, their county is larger than average. The average county has 98,073 residents, whereas the average person lives in a county of size 1,095,564. This is a striking example of "why your classes are larger than average." In

Sampling procedure	Exp. % of blacks in the country	Exp. # of blacks in the county
(A) Aggregate Sample a random county	8.9%	12,293
(C1) Perceived (white) Sample a random white person	10.5%	114,779
(B) Overall Sample a random person	12.5%	148,213
(C2) Perceived (black) Sample a random black person	25.7%	251,006

TABLE 1: The percentage and number of people who identify as 'black alone' in the 2010 US Census, broken down by county. When measured as a proportion, the *ordering* of sampling procedures (B), (C1) and (C2) will always be the same. When measured as a cardinality, the ordering of sampling procedure (A) must be less than that of (C2).

this case, the two different sampling procedures yield a more than ten-fold difference in "average" county size!

Next, we consider diversity as measured by these different sampling procedures.

If we focus on the population that identifies themselves as "black alone," we find large discrepancies in measures of diversity depending on who you ask (or how you sample).

In the US population in 2010, 12.5% of the population identified as "black alone," but that does not mean that an average person experiences a racial mix that is 12.5% black. In fact the average county is less than 9% black, whereas the average white person lives in a county that is more than 10% black, and the average black person lives in a county that is more than 25% black.

If we measure diversity in terms of the absolute number of individuals (rather than percentages), then we find similar discrepancies. The average county has 98,073 residents, 12,293 of whom are black, and yet the average white person lives in a county with 148,213 blacks (and the average black person lives in a county with over 250,000 blacks). These data are summarized in Table 1.

Table 1 highlights a fundamental problem with discussing "diversity" as if it were a single concept. For any given group (e.g., those who identify as "black alone"), the perceived diversity will always find them living with a higher percentage of people from their group than the overall diversity (as measured by asking a random person, or looking at the overall percentage).

It is important to note that this discrepancy in diversity measures will always exist unless every county has exactly the same ratio of populations. In particular, even if every person chose their county randomly, we would still find that a random person lives in a county that is less black than a random black person. See the supplemental materials for an analysis of this situation of random assignment.

School districts

We now examine how this diversity paradox manifests itself in Philadelphia area schools. The data were provided by the School District of Philadelphia through the OpenDataPhilly project [**10**].

Sampling procedure	Exp. % of blacks in the school	Exp. # of blacks in the school
(A) Aggregate Sample a random school	54.3%	310
(C1) Perceived (white) Sample a random white person	43.4%	266
(B) Overall Sample a random person	79.1%	400
(C2) Perceived (black) Sample a random black person	88.5%	435

TABLE 2: The percentage and number of black students in the Philadelphia public schools in 2012-–2013. The table highlights four different sampling procedures for identifying a school. (A) Aggregate: choose a random school and look at the diversity in that school. (B) Overall diversity: choose a random student and look at the school he or she attends. (C1) Choose a random white student and look at the school he or she attends. (C2) Choose a random black person and look at the school he or she attends. The crucial observation is that when measured as a proportion, the *ordering* of sampling procedures (B), (C1), and (C2) will always be the same. When measured as a cardinality, it must be that the expected number of blacks in a classroom in terms of Aggregate Diversity (A), will always be smaller than the expected number of blacks in a classroom from the perspective of a black student (C2).

We examined the enrollment in Philadelphia public schools during the academic year 2012-–2013. During this time, there were a total of 137,020 students, enrolled in 240 distinc schools, meaning the "average" school had 570 students.

Table 2 summarizes the proportions of black students and white students using different metrics for diversity. In these data, 79% of all students were black, so white students were in the minority. In this situation, we see that the average black person is in a school that is 88.5% black (more than we might expect from the overall statistic).

Similarly, although the student population is only 21% white, the average white person is in a school that is 57% white, or 43% black (compared to the overall 79% of students who are black.)

Both whites and blacks experience schools that are much more like them than the simple aggregate statistics (79% black, 21% white) would indicate. Consistent with the results we saw in the census data, there is larger effect—in terms of absolute percentages—on the minority (whites in this case) than on the majority (blacks).

The discrepancy between measures of diversity

In this section, we examine the mathematics behind the diversity discrepancy. The discrepancy is systematic. We first show that when diversity is measured as a proportion, the aggregate diversity (A) and overall diversity (B) can be no greater than the perceived diversity (C). Then we show that when diversity is measured by cardinality, the aggregate diversity (A) can be no greater than the perceived diversity (C).

Measuring diversity as a proportion Suppose a school has N_f females and N_m males, divided into c classrooms, and each student is in exactly one class. Let f_i denote the number of females in classroom i, and let m_i denote the number of males.

Then the *overall diversity* is $\frac{N_f}{N_f+N_m}$. In other words, if you choose a random person from the school, and look at his or her classroom, it will, on average, have a fraction $\frac{N_f}{N_f+N_m}$ of females.

This is different from the *perceived* gender diversity. If we ask a random male about the composition of his class, then he will (on average) experience a higher proportion of males, and if we ask a random female, then she will experience a larger proportion of females.

If we sample a student at random from the population, then the probability the student is in class i is $\frac{f_i+m_i}{N_f+N_m}$, and conditioned on this event, the proportion of females in the class is $\frac{f_i}{f_i+m_i}$, thus the expected proportion is

$$\sum_{i=1}^{c} \frac{f_i+m_i}{N_f+N_m} \cdot \frac{f_i}{f_i+m_i} = \frac{N_f}{N_f+N_m}$$

On the other hand, if we pick a random female, we have an $\frac{f_i}{N_f}$ chance that she will be in classroom i, where the proportion of females is $\frac{f_i}{f_i+m_i}$. Thus, the expected proportion of females in her class is

$$\sum_{i=1}^{c} \frac{f_i}{N_f} \cdot \frac{f_i}{f_i+m_i} = \frac{1}{N_f} \sum_{i=1}^{c} \frac{f_i^2}{f_i+m_i}$$

Lemma 1 (The overall diversity is always lower than the perceived diversity)**.** *Let N_f and N_m be the number of females and males in the population, respectively. Let f_i and m_i be the number of females and males in classroom i, respectively. Then*

$$\frac{N_f}{N_f+N_m} \leq \frac{1}{N_f} \sum_{i=1}^{c} \frac{f_i^2}{f_i+m_i}$$

Proof of Lemma 1. The proof of Lemma 1 follows from a direct application of Jensen's inequality, which states that if φ is a convex function and $\{a_i\}$, $\{x_i\}$ are nonnegative real numbers then

$$\varphi\left(\frac{\sum_i a_i x_i}{\sum_i a_i}\right) \leq \frac{\sum_i a_i \varphi(x_i)}{\sum_i a_i},$$

with equality if and only if all the x_i are equal or φ is linear.

Setting $\varphi(x) \stackrel{\text{def}}{=} \frac{1}{x}$, and $a_i \stackrel{\text{def}}{=} f_i$, and $x_i \stackrel{\text{def}}{=} \frac{f_i+m_i}{f_i}$, Jensen's inequality states that

$$\varphi\left(\frac{\sum_i a_i x_i}{\sum_i a_i}\right) \leq \frac{\sum_i a_i \varphi(x_i)}{\sum_i a_i}.$$

By definition of $\varphi(\cdot)$, this means

$$\frac{\sum_i a_i}{\sum_i a_i x_i} \leq \frac{\sum_i \frac{a_i}{x_i}}{\sum_i a_i}.$$

Cross-multiplying yields

$$\frac{\left(\sum_i a_i\right)^2}{\sum_i a_i x_i} \leq \sum_i \frac{a_i}{x_i}.$$

By definition of a_i and x_i, this gives

$$\frac{\left(\sum_i f_i\right)^2}{\sum_i (f_i + m_i)} \leq \sum_i \frac{f_i^2}{f_i + m_i},$$

which yields the desired result

$$\frac{N_f^2}{N_f + N_m} \leq \sum_i \frac{f_i^2}{f_i + m_i}$$

■

In other words, Lemma 1 states that if we choose a woman from the population, the expected proportion of females in her class will be higher than if we chose a random person from the population. Similarly, if we choose a random male from the population, the expected proportion of males in his class will be higher than if we chose a random person from the population. Every group experiences classrooms whose populations are "more like them" than you might expect.

Next, we compare the *aggregate* diversity and the *perceived* diversity. The aggregate diversity measures the expected proportion of females in a random *classroom* whereas the perceived diversity measures the expected proportion experienced by a randomly chosen female. If a classroom is chosen uniformly at random, then classroom i has a $\frac{1}{c}$ chance of being chosen, and the proportion of females in that classroom is $\frac{f_i}{f_i+m_i}$. Thus the aggregate diversity is $\frac{1}{c}\sum_i \frac{f_i}{f_i+m_i}$. The aggregate diversity corresponds to commonly published metrics (e.g. the diversity of a randomly chosen college [**8**]). Lemma 2 shows that the aggregate diversity will always be lower than the perceived diversity.

Lemma 2 (The aggregate diversity is always lower than the perceived diversity when measured by proportion)**.**

$$\frac{1}{c}\sum_i \frac{f_i}{f_i + m_i} \leq \frac{1}{N_f}\sum_i \frac{f_i^2}{f_i + m_i}.$$

Proof. First, note that if $x \in \mathbb{R}^c$, then $\|x\|_1^2 \leq c\|x\|_2^2$. Since $\sum_i f_i = N_f$, we have

$$\frac{N_f^2}{c} = \frac{\left(\sum_i f_i\right)^2}{c} \leq \sum_i f_i^2.$$

This implies

$$0 \leq \left(\sum_i f_i^2\right) - \frac{N_f^2}{c}.$$

Grouping terms, this means

$$0 \leq \sum_i f_i \left(f_i - \frac{N_f}{c}\right).$$

Dividing both sides by $\sum_i f_i + m_i$ yields

$$0 \leq \frac{\sum_i f_i \left(f_i - \frac{N_f}{c}\right)}{\sum_i f_i + m_i}.$$

Now, note that whenever the b_i are nonnegative

$$\frac{\sum_i a_i}{\sum_i b_i} = \sum_i \frac{a_i}{\sum_j b_j} \leq \sum_i \frac{a_i}{b_i}.$$

Thus, we conclude that

$$0 \leq \frac{\sum_i f_i \left(f_i - \frac{N_f}{c}\right)}{\sum_i f_i + m_i} \leq \sum_i \frac{f_i \left(f_i - \frac{N_f}{c}\right)}{f_i + m_i},$$

which means

$$\frac{N_f}{c} \sum_i \frac{f_i}{f_i + m_i} \leq \sum_i \frac{f_i^2}{f_i + m_i}.$$

■

Remark 1. The overall diversity $\left(\frac{N_f}{N_f + N_m}\right)$ and the aggregate diversity $\left(\frac{1}{c}\sum_i \frac{f_i}{f_i + m_i}\right)$ are incomparable. In particular, there exist partitions of students (i.e., positive integers, $\{f_i\}$ and $\{m_i\}$) for which

$$\frac{N_f}{N_f + N_m} > \frac{1}{c} \sum_i \frac{f_i}{f_i + m_i}$$

and also partitions for which

$$\frac{N_f}{N_f + N_m} < \frac{1}{c} \sum_i \frac{f_i}{f_i + m_i}$$

This means that neither of Lemmas 1 and 2 strictly dominates the other. For example, if there are two classes, the first with a 0 females and 3 males, and the second with 2 females and 3 males, then the aggregate diversity is $1/5$, and the overall diversity is $1/4$. On the other hand, if there are two classes, the first with 3 females and 0 males, and the second with 1 female and 4 males then the aggregate diversity is $3/5$, and the overall diversity is $1/2$.

Measuring diversity by cardinality

Now we examine the relationship between aggregate and perceived diversity, when diversity is measured by cardinality, i.e., the absolute number of minority students in a class (rather than the proportion of minority students).

We begin by reviewing some standard notation. For a discrete set of real numbers $\{x_i\}_{i=1}^N$, we use the notation

$$E(\{x_i\}) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N x_i$$

to denote the mean (expectation of a randomly chosen x_i) and $\text{Var}(\{x_i\})$ to denote the population variance

$$\text{Var}\left(\{x_i\}_{i=1}^N\right) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N (x_i - E(\{x_i\}))^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - (E(\{x_i\}))^2$$

As usual, this means that $\text{Var}(\{x_i\}) \geq 0$.

Lemma 3 (The aggregate diversity is always lower than the perceived diversity when measured by cardinality)**.** *If there are N_f female students, and N_m male students divided into c classrooms, where classroom i has f_i female students and m_i male students, then a randomly chosen classroom will (on average) have N_f/c females, and a randomly chosen female student will, on average, be in a class with $\frac{N_f}{c} + \frac{c}{N_f}\text{Var}(\{f_i\})$ female students.*

Proof. There are N_f female students, and c classrooms, so a randomly chosen *classroom* will, on average, have $\frac{N_f}{c}$ females.

If we choose a random female from the school, then she will be in class i with probability $\frac{f_i}{N_f}$, and in that case, her class will have f_i female members. Thus on average, a randomly chosen class will have

$$\frac{1}{N_f}\sum_{i=1}^{c} f_i^2$$

female members. Simplifying, we find

$$\begin{aligned}
\frac{1}{N_f}\sum_{i=1}^{c} f_i^2 &= \frac{c}{N_f}\left(\frac{1}{c}\sum_i f_i^2\right) \\
&= \frac{c}{N_f}\left(\frac{1}{c}\sum_i f_i^2 - E(\{f_i\})^2 + E(\{f_i\})^2\right) \\
&= \frac{c}{N_f}\left(\frac{1}{c}\sum_i f_i^2 - E(\{f_i\})^2\right) + \frac{c}{N_f}E(\{f_i\})^2 \\
&= \frac{c}{N_f}\text{Var}(\{f_i\}) + \frac{c}{N_f}E(\{f_i\})^2 \\
&= \frac{c}{N_f}\text{Var}(\{f_i\}) + \frac{N_f}{c}
\end{aligned}$$

Thus the average woman is in a class with $\frac{N_f}{c} + \left(\frac{c}{N_f}\text{Var}(\{f_i\})\right)$women. When all the classes have exactly the same number of women (i.e., all the f_i are the same) then the variance is zero, and the perceived diversity is the same as aggregate diversity. Since variances are always nonnegative, we conclude that sampling procedure (A) can never be greater than sampling procedure (C). ■

Note that Lemma 3 follows from essentially the same calculation that explains why your classes are larger than average [**6**].

Remark 2. In absolute terms, the perceived and overall diversity are incomparable. In other words, the expected number of females in a random student's class can be higher or lower than the expected number of females in a randomly chosen female's class.

This means that when diversity is measured by cardinality, the overall diversity need not be smaller than the perceived diversity. This is in contrast to Lemma 2 which states that when diversity is measured as a proportion, the overall diversity is always smaller than the perceived diversity.

To see why this can happen when diversity is measured by cardinality, imagine a school with ten female and ten male students divided into two classrooms of size ten. Suppose the first classroom has nine females one male, while the second has nine males and one female. The aggregate diversity is 5, the perceived diversity is 8.2 and the overall diversity is 5. Imagine that more male students begin to join the first class (which contains nine females). As the number of male students joining the first class increases the aggregate diversity and the perceived diversity will remain the same, but the overall diversity will increase towards nine, since a randomly sampled student will become more and more likely to be in the first classroom containing nine females.

Conclusion

This note shows that (on average) individuals will find that their peers are more like them than one might expect by examining overall statistics. In other words, the diversity perceived by individuals will be less than the reported aggregate diversity. This diversity discrepancy is not uniformly good or bad, but must be taken into account when discussing questions related to diversity.

One consequence of this observation is that minorities experience a world that is slightly less biased than the raw statistics might indicate. For example, the lack of women in STEM fields is a widely acknowledged problem and in 2009, 24% of the STEM workforce was female [**2**]. The masculine culture of male-dominated STEM fields has been put forward as a leading cause of this disparity [**3**]. This note indicates that a randomly chosen woman in a STEM field, will (on average) find that more than 24% of her colleagues are female.

This note has a negative connotation for the majority group. Members of the majority experience a world that is even more skewed than the raw numbers indicate. Returning to the example above, a randomly chosen male in a STEM field will have *fewer* than 24% female colleagues.

If people are randomly assigned to classrooms, teams or neighborhoods, then the discrepancy between aggregate diversity and perceived diversity is usually small enough to be ignored (see the supplemental materials). In the real world, where a variety of social, cultural and economic factors cause groups to cluster together, the difference between the overall and perceived diversity can be extremely large. This discrepancy holds whether you measure diversity in absolute numbers or as a proportion.

The key take-home here is that the overall diversity statistics we see may be very different from the diversity that individuals *experience*, and this discrepancy needs to be taken into account when we are constructing teams, schools and neighborhoods.

REFERENCES

[1] Almquist, Z. W. (2010). US census spatial and demographic data in R: the US census 2000 suit of packages. *J. Statist. Softw*. 37(6): 1–31. 10.18637/jss.v037i06

[2] Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., Doms, M. E. (2011). Women in STEM: A gender gap to innovation. ESA Issue Brief 04-11, US Department of Commerce, 2011.

[3] Cheryan, S., Ziegler, S. A., Montoya, A. K., Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychol. Bullet*. 143(1): 1–35. doi.org/10.1037/bul0000052

[4] Field, S. L. (1991). Why your friends have more friends than you do. *Amer. J. Sociol*. 96(6): 1464–1477. doi.org/10.1086/229693

[5] Good, I. J. (1973). Is the size of our galaxy surprising? *Amer. Statist*. 27(1): 42–43. doi.org/10.1080/00031305.1973.10478977

[6] Hemenway, D. (1982). Why your classes are larger than "average". *Math. Mag.* 55(3): 162–164. doi.org/10.1080/0025570X.1982.11976974

[7] Hodas, N. O., Kooti, F., Lerman, K. (2013). Friendship paradox redux: Your friends are more interesting than you. In: *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*. https://www.aaai.org/Library/ICWSM/icwsm13contents.php.
[8] O'Leary, B., Hatch, J. (2016). Student diversity at more than 4600 institutions. *The Chronicle of Higher Education*. https://www.chronicle.com/article/student-diversity-at-4-725-institutions
[9] University of Pennsylvania: Facts and Figures. http://diversity.upenn.edu/diversity_at_penn/facts_figures.
[10] Open data philly school district school information. (2017). http://webgui.phila.k12.pa.us/offices/o/open-data-initiative/documents/sdp_school_20130510.zip.

Summary. Diversity has many potential benefits, and increasing diversity in the classroom and the workplace is a high-priority goal for managers and policy-makers. Although there is nearly uniform support for increasing diversity, there is less agreement about how to quantify diversity. This note observes that there is a systematic bias between how diversity is perceived by an individual and the natural measures of population diversity. In particular, most individuals will find themselves among people who are "more like them" than the aggregate statistics indicate. For example, considering racial diversity in the classroom, do you measure diversity as the proportion of minority students in the average class, or the average proportion as experienced by a member of the minority? Using examples from the US Census and the Philadelphia Public Schools we show that these two metrics can be quite different, and we give a general mathematical treatment of this discrepancy.

BRETT HEMENWAY (MR Author ID: 863162), Ph.D. is a research assistant professor in the Department of Computer and Information Science at the University of Pennsylvania, where his research focuses on coding theory, cryptography, and mathematical modeling.

DAVID HEMENWAY (MR Author ID: 1201712), Ph.D. is a Professor of Health Policy at the Harvard TH Chan School of Public Health and Director of the Harvard Injury Control Research Center. He is delighted to be writing mathematical articles with his son.

A Direct Proof of Pitot's Theorem

QUANG HUNG TRAN
High School for Gifted Students
Hanoi University of Science
Vietnam National University at Hanoi
tranquanghung@hus.edu.vn

Theorem 1 (Pitot, 1725)**.** *A convex quadrilateral $ABCD$ can have an inscribed circle if and only if its sides satisfy*

$$AB + CD = BC + AD. \tag{1}$$

Several proofs of this result are given by Josefsson [**1**]. We provide a novel proof based on congruent triangles.

It is not difficult to prove necessity by using the equality of the lengths of the two tangents to a circle from a fixed point outside the circle. So, we will only give the proof of sufficiency.

Proof. If a convex quadrilateral $ABCD$ satisfies equation (1), then we shall prove that it has an inscribed circle.

We assume that the bisectors of angles $\angle ADC$ and $\angle BCD$ meet at point I, and that the angle bisectors of angles $\angle DAB$ and $\angle ADC$ meet at point J. (See Figure 1).

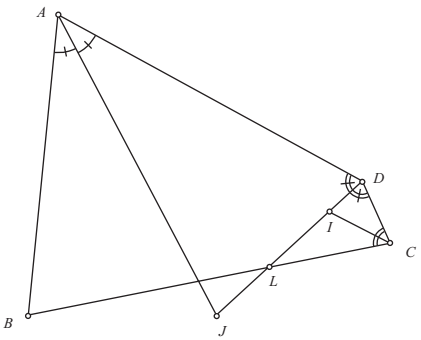


Figure 1 The quadrilateral $ABCD$ with three of its angle bisectors. Note that point I can be assumed to lie inside the quadrilateral.

Since $ABCD$ is convex, if J does not lie inside $ABCD$, then J and D are not on the same side with respect to line BC. Therefore, segment DJ must meet side BC at a point L. Since CI is the bisector of $\angle LCD$, I must lie on segment LD, implying that I lies inside $ABCD$. If J lies inside $ABCD$, then we will use J instead of I because, from equation (1), we see that the roles of the opposite sides of quadrilateral $ABCD$ are the same. Hence, without loss of generality, we can assume that I lies inside $ABCD$.

Since I is the intersection point of the bisectors of angles $\angle ADC$ and $\angle BCD$, it follows that I is equidistant from sides AD, DC, and CB. Therefore, there exists a circle ω (with center I) that is tangent to sides AD, DC, and CB. Let E, F, and G denote the points of tangency of ω with, respectively, sides BC, AD, and CD. See Figure 2.

Math. Mag. **95** (2022) 59–60. doi:10.1080/0025570X.2022.2001254

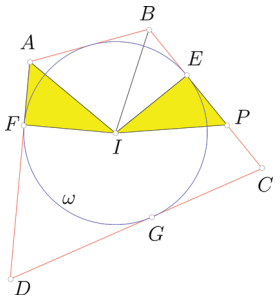


Figure 2 Segments AD, BC, and CD are tangent to the circle ω. We only need to show that AB is also tangent to ω.

Since the lengths of the two tangent line segments from a point outside a circle are equal, we have

$$DG = DF \quad \text{and} \quad CG = CE. \tag{2}$$

From (1) and (2), we deduce that

$$AB = AF + BE. \tag{3}$$

Let P be the point on the extension of ray BE such that

$$EP = AF. \tag{4}$$

We also have that $IE = IF$ because both are radii of ω, and that $\angle IFA$ and $\angle IEP$ are both right. Therefore, $\triangle IFA$ and $\triangle IEP$ are congruent by SAS. We therefore obtain

$$AI = IP. \tag{5}$$

From (3) and (4), we also have

$$AB = BP. \tag{6}$$

From (5) and (6), we get two triangles ABI and PBI are congruent by SSS. Therefore, the distance from I to AB is equal to IE, which is the radius of ω. This implies that AB is tangent to ω. This completes the proof. ■

Acknowledgments The author wishes to thank Professor Jean-Louis Ayme from France and Professor Martin Josefsson from Sweden for proofreading this work. The author would also like to thank the referee for careful reading and valuable comments.

REFERENCES

[1] Josefsson M. (2019). On Pitot's theorem, *Math. Gaz.*. 103(557): 333–337. doi.org/10.1017/mag.2019.70

Summary. Pitot's theorem provides necessary and sufficient conditions for a convex quadrilateral to have an inscribed circle. There are many known proofs of this result. This paper present a new and direct proof of Pitot's theorem using congruent triangles.

QUANG HUNG TRAN received a BS in pure mathematics in 2008, and a Master's degree in pure mathematics in 2011, from the Vietnam National University at Hanoi. He is a geometry teacher at the High School for Gifted Students (HSGS) of Vietnam National University at Hanoi. His main interests include many aspects of geometry. Other than teaching and doing mathematics, he enjoys reading and spending time with his family.

Report on the 50th Annual USA Mathematical Olympiad

BÉLA BAJNOK
Gettysburg College
Gettysburg, PA 17325
bbajnok@gettysburg.edu

The USA Mathematical Olympiad (USAMO) is the final round in the American Mathematics Competitions series for high school students, organized each year by the Mathematical Association of America. The competition follows the style of the International Mathematics Olympiad. It consists of three problems each on two consecutive days, with an allowed time of four-and-a-half hours on both days.

The 50th annual USAMO was given on Tuesday, April 13, 2021 and Wednesday, April 14, 2021. This year, 288 students were invited to take the USAMO and, as in 2020, the competition was administered online. The names of the winners and honorable mentions, as well as more information on the American Mathematics Competitions program, can be found at the site https://www.maa.org/math-competitions. Below we present the problems and solutions of the competition; a similar article for the USA Junior Mathematical Olympiad (USAJMO), offered to students in grade 10 or below, can be found in the January 2022 issue of the *College Mathematics Journal*.

The problems of the USAMO are chosen—from a large collection of proposals submitted for this purpose—by the USAMO/USAJMO Editorial Board, whose co-editors-in-chief this year were Evan Chen and Jennifer Iglesias, with associate editors Ankan Bhattacharya, John Berman, Zuming Feng, Sherry Gong, Alison Miller, Maria Monks Gillespie, and Alex Zhai. This year's problems were created by Ankan Bhattacharya, Mohsen Jamaali, Shaunak Kishore, Carl Schildkraut, Zoran Sunic, and Alex Zhai.

The solutions presented here are those of the present author, relying in part on the submissions of the problem authors and members of the editorial board. Each problem was worth 7 points; the nine-tuple $(n; a_7, a_6, a_5, a_4, a_3, a_2, a_1, a_0)$ states the number of students who submitted a article for the relevant problem, followed by the numbers who scored $7, 6, \ldots, 0$ points, respectively.

Problem 1 (226; 140, 0, 0, 0, 1, 0, 11, 74); *proposed by Ankan Bhattacharya*. Rectangles BCC_1B_2, CAA_1C_2, and ABB_1A_2 are erected outside an acute triangle ABC. Suppose that

$$\angle BC_1C + \angle CA_1A + \angle AB_1B = 180^\circ.$$

Prove that lines B_1C_2, C_1A_2, and A_1B_2 are concurrent.

Solution. Let ω_A, ω_B, and ω_C be the circumcircles of rectangles BCC_1B_2, CAA_1C_2, and ABB_1A_2, respectively. Define P to be the foot of the altitude from A to B_1C_2.

Observe that, since $\angle ACC_2$ and $\angle APC_2$ are both 90°, P lies on ω_B by the inscribed angle theorem and, therefore, $\angle APC$ and $\angle CA_1A$ are supplementary angles. A similar argument shows that P lies on ω_C and thus $\angle APB$ and $\angle AB_1B$ are supplementary angles as well.

Math. Mag. **95** (2022) 61–70. doi:10.1080/0025570X.2022.2001226
MSC: 00A05

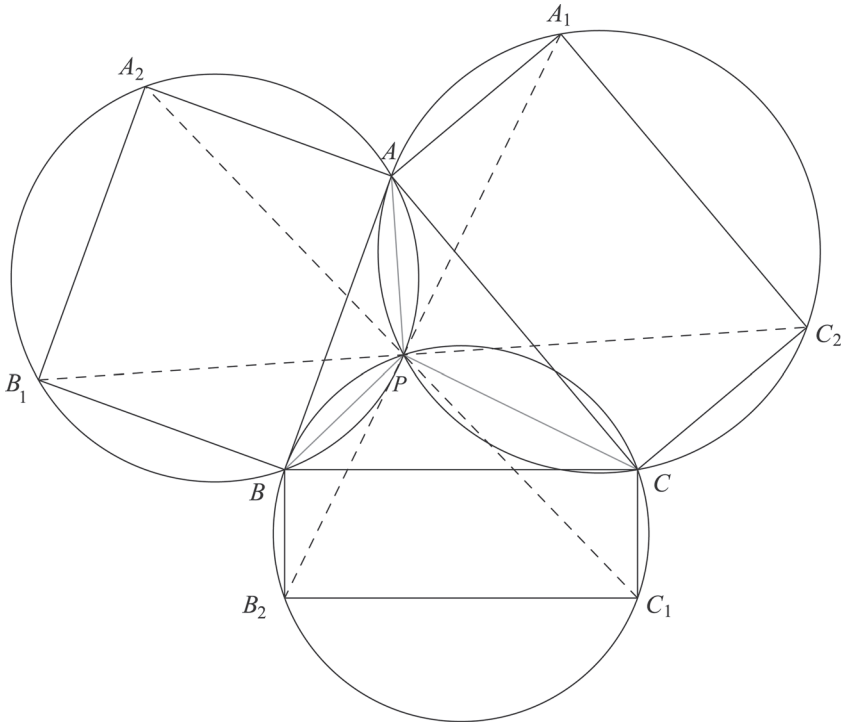


Figure 1 The main circles used in Problem 1.

But then

$$\angle BPC = 360^\circ - (\angle APC + \angle APB) = \angle CA_1A + \angle AB_1B,$$

which, by the given equation, yields

$$\angle BPC = 180^\circ - \angle BC_1C,$$

and thus P lies on ω_A as well. Therefore, P is the (unique) common point of all three circles.

Similar arguments would prove that the feet of the altitudes from B and C to C_1A_2 and A_1B_2, respectively, are on each of the three circles, and thus must coincide with P. But then lines B_1C_2, C_1A_2, and A_1B_2 are concurrent, as claimed.

Problem 2 (199; 76, 9, 1, 4, 13, 18, 1, 77); *proposed by Zoran Sunic*. The Planar National Park is a subset of the Euclidean plane consisting of several trails which meet at junctions. Every trail has its two endpoints at two different junctions, whereas each junction is the endpoint of exactly three trails. Trails only intersect at junctions (in particular, trails only meet at endpoints). Finally, no trails begin and end at the same two junctions.

A visitor walks through the park as follows: she begins at a junction and starts walking along a trail. At the end of that first trail, she enters a junction and turns left. On the next junction she turns right, and so on, alternating left and right turns at each junction. She does this until she gets back to the junction where she started. What is the largest possible number of times she could have entered any junction during her walk, over all possible layouts of the park?

One possible layout for the park is shown in Figure 2.

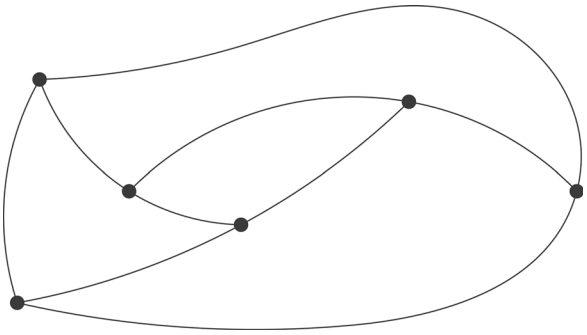


Figure 2 One possible layout of the park, with six junctions and nine trails.

Solution. The answer is three times. We begin by exhibiting an example of a park layout which features three visits. Sketched in Figure 3 is one of many possible constructions. The path starts from C, walks toward A, and continues as follows:

$$C \to A \to H \to I \to F \to G \to D \to B$$
$$\to A \to H \to E \to F \to G \to J \to B \to A \to C$$

As we see, this path visits A three times.

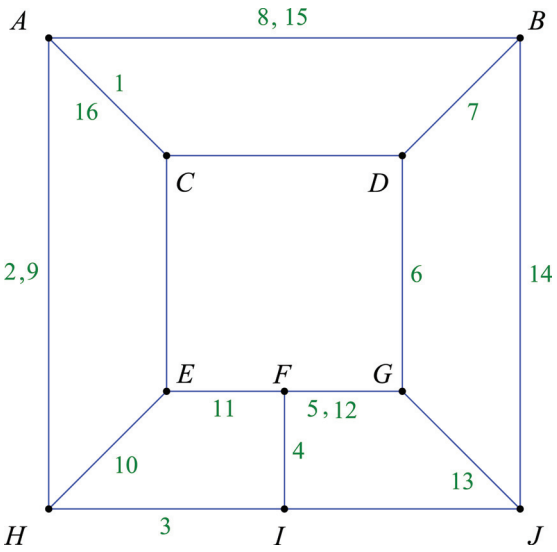


Figure 3 An example achieving three visits.

We will prove that the visitor cannot visit any junction more than three times. (This trivially holds fore the initial/terminal junction.) Note that if a junction were to be visited four times or more, then this would mean four or more arrivals and four or more departures. This implies that one of the three trails meeting at that junction had to have been traversed at least three times. Therefore, it suffices to show that no trail can be on the visitor's path more than twice.

Suppose that there is a trail that the visitor walked on three or more times. (Note that this trail cannot be adjacent to the junction where her walk started.) This implies that at least two of those times she turned in the same direction (left or right) when she reached the end of the trail. Letus assume that the mth and nth trail during her walk is the same for some $2 \leq m < n$ with the same turn at the end; we may further assume that this is the trail with the smallest possible m. Let A and B denote the junctions at the two ends of this trail.

Now if she walked along this trail both times in the same direction, say from A to B, and made the same turn at the end (e.g., left), then her $(m-1)$st and $(n-1)$st trails were also the same, and she made the same turn when she got to A (right). This contradicts the minimality of m. On the other hand, if once she walked from A to B and then later from B to A, but turning in the same direction at the end both times (e.g., left), then her $(m-1)$st trail was the same as her $(n+1)$st, and she made the same turn at the ends of these two trails as well (right), again contradicting the minimality of m. This completes our proof.

Problem 3 (179; 7, 4, 1, 0, 1, 4, 107, 55); *proposed by Shaunak Kishore and Alex Zhai*. Let $n \geq 2$ be an integer. An $n \times n$ board is initially empty. Each minute, you may perform one of three moves:

- If there is an L-shaped tromino region of three cells without stones on the board (see Figure 4; rotations not allowed), you may place a stone in each of those cells.

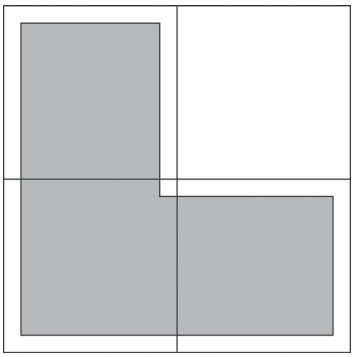


Figure 4 Three cells forming an L-shaped tromino.

- If all cells in a column have a stone, then you may remove all stones from that column.
- If all cells in a row have a stone, then you may remove all stones from that row.

For which n is it possible that, after some nonzero number of moves, the board has no stones?

Solution. We claim that the answer is all multiples of 3. First, we show that the procedure is possible in each such case.

When n is divisible by 3, one may divide the board into 3×3 sub-squares; for brevity, let us refer to these $(n/3)^2$ sub-squares as *cages*. We then follow the procedure illustrated in Figure 5, as follows.

- First, we put two nonoverlapping L-trominoes in each cage, as shown in the first step.

- This causes every center column of each cell to be completely filled. Thus, we may remove all $n/3$ columns which correspond to the center columns of cages, as shown in the second step.
- In each cage, we then place one L-tromino as shown in the third step.
- Now the board consists of $2n/3$ completely filled rows, so we may eliminate them all.

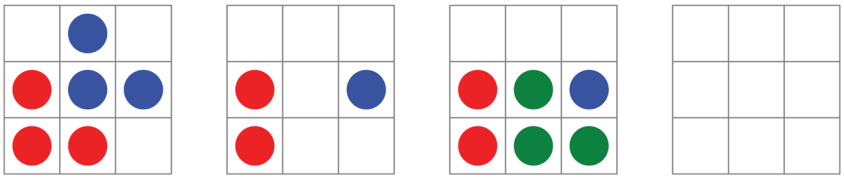


Figure 5 The four-step procedure that clears all stones.

We now prove that if after some sequence of moves no stones remain, then n must be a multiple of 3. We will employ what is usually called the *polynomial method*. Specifically, we make use of the following.

Lemma 1. *Consider the polynomial*

$$f(x, y) = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} d_{i,j} x^i y^j,$$

where the coefficients $d_{i,j}$ are real numbers and $d_{n_1,n_2} \neq 0$. If A_1 and A_2 are sets of real numbers with $|A_1| > n_1$ and $|A_2| > n_2$, then there are elements $a_1 \in A_1$ and $a_2 \in A_2$ for which $f(a_1, a_2) \neq 0$.[*]

Let us introduce some notation. We parametrize the cells of the board by letting (i, j) denote the position of the cell in the ith column (counting from the left) and the jth row (counting from the bottom). We then associate each state of the board with the polynomial

$$A(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{i,j} x^{i-1} y^{j-1},$$

where $c_{i,j}$ is 1 when there is a stone in cell (i, j) and 0 otherwise. This allows us to think of the chain of moves as the sequence $(A_k(x, y))_{k=0}^{m}$ where $A_0(x, y) = 0$ (representing the initial position of the board), $A_m(x, y) = 0$ (expressing the fact that there are no stones on the board after m moves for some $m \in \mathbb{N}$), and where $A_k(x, y)$ results from $A_{k-1}(x, y)$ in one of the following ways:

- $A_k(x, y) = A_{k-1}(x, y) + x^{i-1} y^{j-1}(1 + x + y)$ if a tromino was placed on the board with its lower left corner at position (i, j);
- $A_k(x, y) = A_{k-1}(x, y) - x^{i-1}(1 + y + y^2 + \cdots + y^{n-1})$ if all stones got removed from column i; and
- $A_k(x, y) = A_{k-1}(x, y) - y^{j-1}(1 + x + x^2 + \cdots + x^{n-1})$ if all stones got removed from row j.

[*]Note that this lemma is the two-variable version of the well-known fact that a nonzero polynomial cannot have more roots than its degree. For a simple proof and a variety of applications, see, for example, Section 12.3 in Freud and Gyarmati [1].

We need some more notation. For $i = 1, 2, \ldots, n-1$ and $j = 1, 2, \ldots, n-1$, we let $a_{i,j}$ denote the number of times a tromino was added with its lower-left corner at position (i, j); we then set

$$P(x, y) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} a_{i,j} x^{i-1} y^{j-1}.$$

Furthermore, we set $r(j)$ and $c(i)$ equal to the number of times the jth row and ith column were cleared, respectively, and define $Q(x) = \sum_{i=1}^{n} c_i x^{i-1}$ and $R(y) = \sum_{j=1}^{n} r_j y^{j-1}$. With this notation, the fact that our procedure succeeded can be stated by the equation

$$\begin{aligned} P(x, y)(1 + x + y) - Q(x)(1 + y + y^2 + \cdots + y^{n-1}) \\ - R(y)(1 + x + x^2 + \cdots + x^{n-1}) = 0. \end{aligned}$$

Take A to be the set of nth roots of unity other than 1; that is, the $n-1$ distinct complex numbers a for which

$$\frac{a^n - 1}{a - 1} = 1 + a + a^2 + \cdots + a^{n-1} = 0.$$

Since $P(x, y)$ is a nonzero polynomial with x-degree and y-degree at most $n-2$, the lemma guarantees there are elements $a_1, a_2 \in A$ for which $P(a_1, a_2) \neq 0$. But since substituting $x = a_1$ and $y = a_2$ into our equation yields

$$P(a_1, a_2)(1 + a_1 + a_2) = 0,$$

this can only occur when $1 + a_1 + a_2 = 0$. Therefore, the imaginary parts of a_1 and a_2 are negatives of one another. Recalling that both numbers have norm 1, this implies that their real parts have the same absolute value. But these real parts must then both be negative, and in fact equal to $-\frac{1}{2}$, so a_1 and a_2 are $-\frac{1}{2} \pm \frac{\sqrt{3}}{2} i$. We thus obtain that a_1 and a_2 are cube roots of unity. However, this can only happen if n is divisible by 3. Indeed, if $n = 3q + r$ for some integers q and r with $r \in \{0, 1, 2\}$, then $a_1^r = a_1^n / a_1^{3q} = 1$, which is only possible when $r = 0$, as claimed.

Problem 4 (240; 121, 14, 5, 0, 1, 37, 13, 49); *proposed by Carl Schildkraut*. A finite set S of positive integers has the property that, for each $s \in S$, and each positive integer divisor d of s, there exists a unique element $t \in S$ satisfying $\gcd(s, t) = d$. (The elements s and t could be equal.)

Given this information, find all possible values for the number of elements of S.

Solution. We claim that the possible sizes are 0 and the nonnegative integer powers of 2. Since $S = \emptyset$ and $S = \{1\}$ obviously work, we need to show that a set S of size $n \geq 2$ satisfying the requirements exists if and only if $n = 2^k$ for some positive integer k.

We start by verifying that these values are indeed possible. For a given positive integer k, we construct a set of size 2^k as follows: Suppose that $p_1, q_1, p_2, q_2, \ldots, p_k, q_k$ are $2k$ pairwise distinct positive primes. For an ordered pair of subsets (I, J) of $[k] = \{1, 2, \ldots, k\}$, we will use the notation

$$s(I, J) = \prod_{i \in I} p_i \cdot \prod_{j \in J} q_j.$$

(Recall that the empty product equals 1.) We then consider the set

$$S = \{s(I, J) \mid I \subseteq [k], J = [k] \setminus I\}.$$

Since the elements of S are then in a bijection with the subsets I of $[k]$, we see that $|S| = 2^k$. We need to show that S satisfies the required property.

Given an element $s(I, J)$ of S, we see that its positive divisors are of the form $s(I_0, J_0)$ where $I_0 \subseteq I$ and $J_0 \subseteq J$. Note also that, since I and J form a partition of $[k]$, the sets

$$I' = I_0 \cup (J \setminus J_0) \qquad \text{and} \qquad J' = J_0 \cup (I \setminus I_0)$$

do as well, and thus $s(I', J')$ is an element of S. In fact, it is the unique element of S whose greatest common divisor with $s(I, J)$ equals $s(I_0, J_0)$. Therefore, the set S we constructed satisfies the requirement of the problem.

It remains to be shown that if S is a set satisfying the property and it has size $n \geq 2$, then $n = 2^k$ for some positive integer k. Let $s \geq 2$ be any element of S, and let p be any positive prime divisor of s. We can then write $s = p^e \cdot u$ for some positive integers e and u where u is not divisible by p. We claim that $e = 1$.

Denoting by $d(m)$ the number of positive divisors of a positive integer m, we have $d(s) = (e + 1) \cdot d(u)$. In fact, s has exactly $e \cdot d(u)$ positive divisors that are divisible by p and $d(u)$ that are not. By our assumption, there is an element t of S for which $\gcd(t, s) = p$. Let us assume that $e \geq 2$. We can then see that $t = p \cdot v$ for some positive integer v that is not divisible by p. Furthermore, $d(t) = 2d(v)$, and t has exactly $d(v)$ positive divisors that are divisible by p and also $d(v)$ that are not.

Now according to our requirement for S, the positive divisors of any element s of S are in one-to-one correspondence with the elements of S via the map $a \mapsto \gcd(a, s)$, and thus $|S| = d(s)$. (In particular, all elements of S must have the same number of positive divisors.) Furthermore, for any prime divisor p of s, we have $p|a$ if and only if $p|\gcd(a, s)$. Therefore, S has exactly $e \cdot d(u)$ elements that are divisible by p and $d(u)$ that are not. With the same reasoning, S has exactly $d(v)$ elements that are divisible by p and $d(v)$ that are not. But then $d(u) = d(v)$ and $e \cdot d(u) = d(v)$, from which $e = 1$.

This establishes that each element of S is a product of the same number of pairwise distinct prime numbers. If $s \in S$ is the product of k distinct primes, then $|S| = d(s) = 2^k$. This completes our proof.

Problem 5 (215; 95, 10, 0, 1, 2, 29, 2, 76); *proposed by Mohsen Jamaali*. Let $n \geq 4$ be an integer. Find all positive real solutions to the following system of $2n$ equations:

$$\begin{aligned}
a_1 &= \frac{1}{a_{2n}} + \frac{1}{a_2}, & a_2 &= a_1 + a_3,\\
a_3 &= \frac{1}{a_2} + \frac{1}{a_4}, & a_4 &= a_3 + a_5,\\
a_5 &= \frac{1}{a_4} + \frac{1}{a_6}, & a_6 &= a_5 + a_7,\\
&\vdots & &\vdots\\
a_{2n-1} &= \frac{1}{a_{2n-2}} + \frac{1}{a_{2n}}, & a_{2n} &= a_{2n-1} + a_1.
\end{aligned}$$

First solution. It is easy to see that a solution is provided by $a_1 = a_3 = \cdots = a_{2n-1} = 1$ and $a_2 = a_4 = \cdots = a_{2n} = 2$. We now prove that there are no others.

Taking indices modulo $2n$ and eliminating terms with odd indices, for each $i = 1, 2, \ldots, n$, we have

$$a_{2i} = \frac{1}{a_{2i-2}} + \frac{2}{a_{2i}} + \frac{1}{a_{2i+2}}. \tag{1}$$

Adding up these equations yields

$$\sum_{i=1}^{n} a_{2i} = \sum_{i=1}^{n} \frac{4}{a_{2i}}. \tag{2}$$

According to the harmonic mean–arithmetic mean inequality,

$$\frac{n}{\sum_{i=1}^{n} 1/a_{2i}} \leq \frac{\sum_{i=1}^{n} a_{2i}}{n}, \tag{3}$$

so by (2) we get $\sum_{i=1}^{n} a_{2i} \geq 2n$.

Dividing both sides of (1) by a_{2i} yields

$$1 = \frac{1}{a_{2i-2}a_{2i}} + \frac{2}{a_{2i}^2} + \frac{1}{a_{2i}a_{2i+2}}, \tag{4}$$

and adding the equations results in

$$n = \sum_{i=1}^{n} \left(\frac{1}{a_{2i}} + \frac{1}{a_{2i+2}} \right)^2. \tag{5}$$

Now we use the quadratic mean–arithmetic mean inequality, which gives

$$\frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{1}{a_{2i}} + \frac{1}{a_{2i+2}} \right)^2 \geq \left(\frac{\sum_{i=1}^{n} \left(\frac{1}{a_{2i}} + \frac{1}{a_{2i+2}} \right)}{n} \right)^2, \tag{6}$$

so by (5) and (2) we get

$$1 \geq \left(\frac{\sum_{i=1}^{n} \frac{2}{a_{2i}}}{n} \right)^2 = \left(\frac{\frac{1}{2} \cdot \sum_{i=1}^{n} a_{2i}}{n} \right)^2$$

and thus $\sum_{i=1}^{n} a_{2i} \leq 2n$.

This means that each of our inequalities is an equality, and therefore $a_{2i} = 2$ for all i. This in turn implies that $a_{2i-1} = 1$ for all i, as claimed.

Second solution. We write $m = \min\{a_{2i} \mid 1 \leq i \leq n\}$ and $M = \max\{a_{2i} \mid 1 \leq i \leq n\}$, and assume that $m = a_{2j}$ and $M = a_{2k}$. Then equation (1) from the first solution yields

$$m = \frac{1}{a_{2j-2}} + \frac{2}{m} + \frac{1}{a_{2j+2}} \geq \frac{1}{M} + \frac{2}{m} + \frac{1}{M}$$

and

$$M = \frac{1}{a_{2k-2}} + \frac{2}{M} + \frac{1}{a_{2k+2}} \leq \frac{1}{m} + \frac{2}{M} + \frac{1}{m}.$$

Therefore,

$$m \geq \frac{2}{m} + \frac{2}{M} \geq M,$$

which can only occur when $m = M$. Therefore, all a_{2i} are equal, from which $a_2 = a_4 = \cdots = a_{2n} = 2$ and $a_1 = a_3 = \cdots = a_{2n-1} = 1$, as claimed.

Problem 6 (133; 17, 2, 1, 1, 3, 1, 5, 103); *proposed by Ankan Bhattacharya*. Let $ABCDEF$ be a convex hexagon satisfying

$$\overline{AB} \parallel \overline{DE}, \qquad \overline{BC} \parallel \overline{EF}, \qquad \overline{CD} \parallel \overline{FA},$$

and

$$AB \cdot DE = BC \cdot EF = CD \cdot FA.$$

Let X, Y, and Z be the midpoints of $\overline{AD}$, $\overline{BE}$, and $\overline{CF}$. Prove that the circumcenter of $\triangle ACE$, the circumcenter of $\triangle BDF$, and the orthocenter of $\triangle XYZ$ are collinear.

Solution. We will prove that the orthocenter of $\triangle XYZ$ is in fact the midpoint of the segment connecting the circumcenter of $\triangle ACE$ and the circumcenter of $\triangle BDF$.

For each pair of adjacent sides of the hexagon, we construct a parallelogram with these two sides. This results in parallelograms $ABCE'$, $BCDF'$, $CDEA'$, $DEFB'$, $EFAC'$, and $FABD'$, as shown in Figure 6. (To aid legibility, Figure 6 is intentionally *not* drawn to scale.)

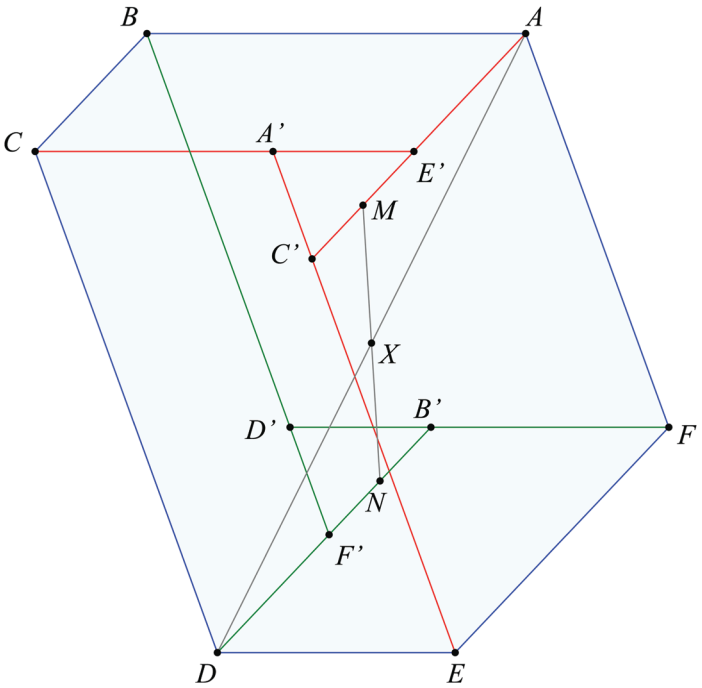


Figure 6 Parallelograms used in the solution to Problem 6.

Note that the assumption that $\triangle XYZ$ is nondegenerate implies that no two opposite sides of the hexagon have the same length; therefore, $\triangle A'C'E'$ and $\triangle B'D'F'$ are also nondegenerate triangles. It is easy to see that these two triangles are translations of one another, as corresponding sides are parallel and have the same length: for example, $\overline{A'E'} \parallel \overline{D'B'}$ and, assuming without loss of generality that $AB > DE$, we have

$$A'E' = CE' - CA' = AB - DE = D'F - B'F = D'B'.$$

Now let M and N be the midpoints of $\overline{C'E'}$ and $\overline{B'F'}$, respectively. Then $MAND$ is a parallelogram, because $\overline{AM}$ and $\overline{DN}$ are parallel, and

$$AM = AE' + \frac{1}{2}E'C' = DF' + \frac{1}{2}F'B' = DN.$$

Therefore, X (the midpoint of $\overline{AD}$) is the midpoint of $\overline{MN}$; similarly, Y is the midpoint of the segment connecting the midpoint of $\overline{A'C'}$ and the midpoint of $\overline{F'D'}$, and Z is the midpoint of the segment connecting the midpoint of $\overline{A'E'}$ and the midpoint of $\overline{B'D'}$. Therefore, the orthocenter of $\triangle XYZ$ is the midpoint of the segment connecting the orthocenters of the medial triangles of $\triangle A'C'E'$ and $\triangle B'D'F'$. Recall that the orthocenter of the medial triangle of a triangle $\triangle$ is the circumcenter of $\triangle$, hence the orthocenter of $\triangle XYZ$ is the midpoint of the segment connecting the circumcenters of $\triangle A'C'E'$ and $\triangle B'D'F'$. We can thus complete our proof by showing that the circumcenters of $\triangle ACE$ and $\triangle A'C'E'$ coincide and that the circumcenters of BDF and $B'D'F'$ coincide. We show the first of these as the second claim can be done similarly.

With ω denoting the circumcircle of $\triangle A'C'E'$, set r equal to the radius of ω, and let d_1, d_2, and d_3 be the distances of A, C, and E from the center of ω, respectively. The power of A to ω is then

$$d_1^2 - r^2 = AE' \cdot AC' = BC \cdot EF;$$

similarly, we have

$$d_2^2 - r^2 = CE' \cdot CA' = AB \cdot DE$$

and

$$d_3^2 - r^2 = EA' \cdot EC' = CD \cdot FA.$$

According to our assumption, the three quantities are the same, which implies that A, C, and E have the same distance from the center of ω, and thus the circumcenters of $\triangle ACE$ and $\triangle A'C'E'$ coincide, as claimed. This completes our proof.

Acknowledgments The author wishes to express his immense gratitude to everyone who contributed to the success of the competition: the students and their teachers, coaches, and proctors; the problem authors; the USAMO Editorial Board; the graders; the Art of Problem Solving; and the AMC Headquarters of the MAA. I am also grateful to Evan Chen for proofreading this article and for providing the figures.

REFERENCE

[1] Freud, R., and Gyarmati, E. (2020). *Number Theory*. Providence, RI, USA: American Mathematical Society.

Summary. We present the problems and solutions to the 50th Annual United States of America Mathematical Olympiad.

BÉLA BAJNOK (MR Author ID: 314851, ORCID 0000-0002-9498-1596) is a professor of mathematics at Gettysburg College and the director of the American Mathematics Competitions program of the MAA.

PROOF WITHOUT WORDS

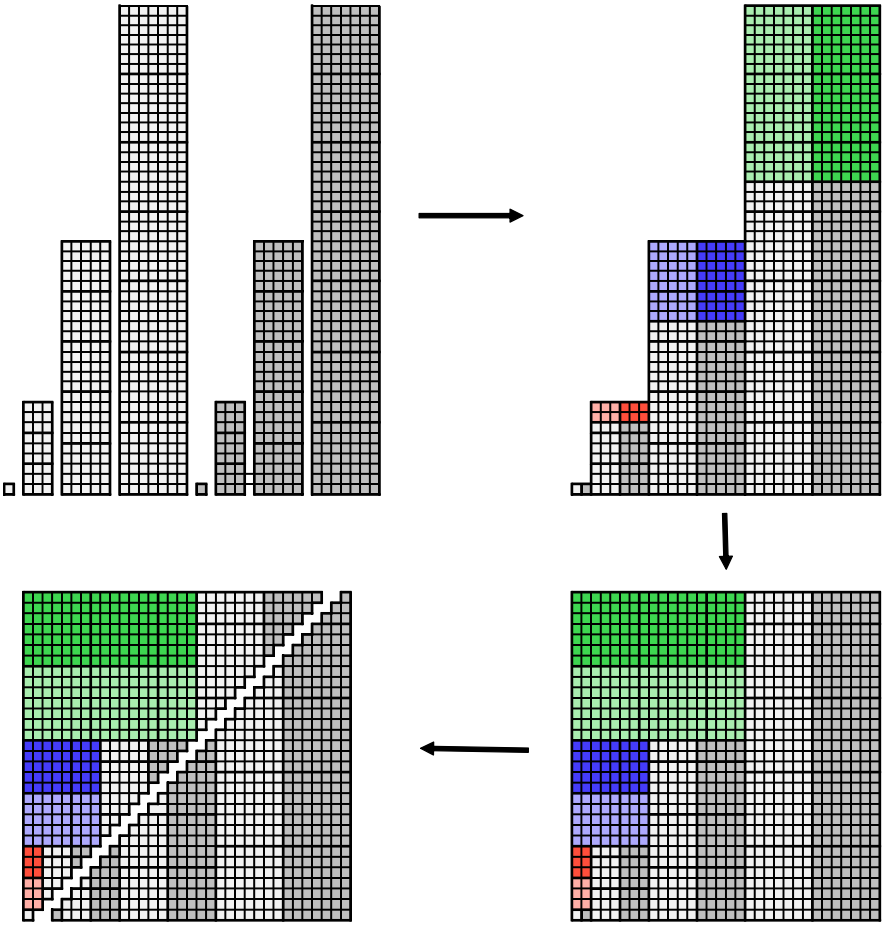
Sums of Odd Cubes Are Triangular Numbers

STEPHAN BERENDONK
Universität Duisburg-Essen
stephan.berendonk@uni-due.de

In 1995, Monte Zerger proved that the sum of the first n odd cubes is equal to the sum of the first $2n^2 - 1$ integers by means of a Proof Without Words. His proof was originally published in this MAGAZINE [2] and was reprinted in Nelsen's anthology of visual proofs [1, p. 91]. The following two proofs of the same fact are similar to each other, but different from Zerger's decomposition.

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

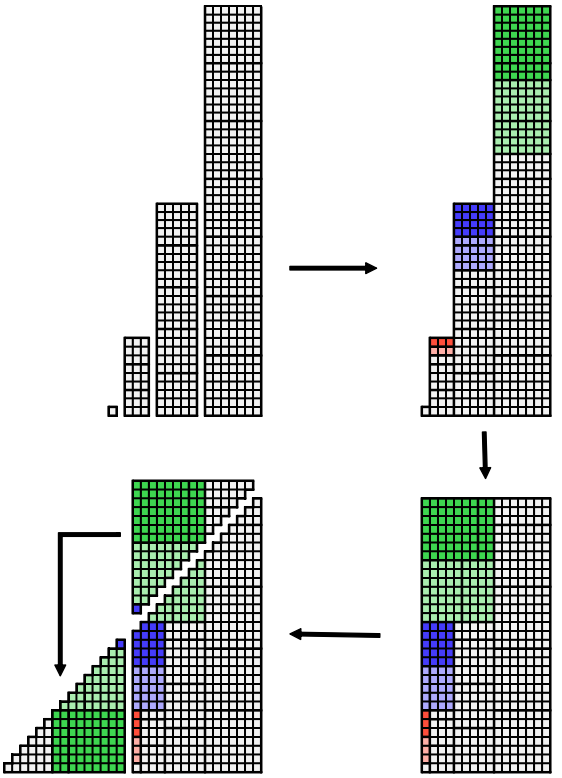
We illustrate this for the case $n = 4$.



*Note that the online version of this article includes color diagrams.

Math. Mag. **95** (2022) 71–72. doi:10.1080/0025570X.2022.2000820

The diagram actually demonstrates that twice the sum of cubes is twice the desired sum, so that ultimately one has to divide both sides by 2. In this regard, the following proof is more direct. We again illustrate this for the case $n = 4$.



Note that one can use analogous proofs to show the sum of the first n *even* cubes is equal to $2(n(n+1))^2$.

REFERENCES

[1] Nelsen, R. B. (1993). *Proofs without Words: Exercise in Visual Thinking*. Washington, DC: Mathematical Association of America.

[2] Zerger, M. J. (1995). The sum of consecutive odd cubes is a triangular number. *Math. Mag.* 68(5): 371. doi.org/10.1080/0025570X.1995.11996357

Summary. Two visual proofs are given for the fact that the sum of the first consecutive odd cubes are triangular numbers.

STEPHAN BERENDONK (MR Author ID: 1055756) received his Ph.D. from the University of Cologne in 2013. He currently works as a postdoc in mathematics education at the University of Duisburg-Essen.

PROBLEMS

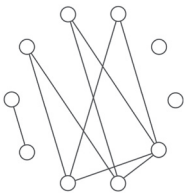
LES REID, *Editor*
Missouri State University

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

RICHARD BELSHOFF, Missouri State University; *MAHYA GHANDEHARI*, University of Delaware; *EYVINDUR ARI PALSSON*, Virginia Tech; *GAIL RATCLIFF*, East Carolina University; *ROGELIO VALDEZ*, Centro de Investigación en Ciencias, UAEM, Mexico; *Assistant Editors*

Correction

Somewhere during the production of the June 2021 issue, an edge was omitted from the graph in Problem 2125, rendering it unsolvable. The correct graph is shown below. We regret the error and thank Elton Bojaxhiu, Enkel Hysnelaj, Didier Pinchon, and



Afshin Ghoreishi for pointing it out.

Proposals

To be considered for publication, solutions should be received by July 1, 2022.

2136. *Proposed by Necdet Batir, Nevşehir HBV University, Nevşehir, Turkey.*

Evaluate

$$\lim_{n\to\infty}\left(\left(\sum_{k=1}^{n}\frac{H_k^2}{k}\right)-\frac{H_n^3}{3}\right),$$

where $H_n=\sum_{k=1}^{n}\frac{1}{k}$ is the nth harmonic number.

Math. Mag. **95** (2022) 73–82. doi:10.1080/0025570X.2022.2008756

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

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

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

2137. *Proposed by the Columbus State University Problem Solving Group, Columbus State University, Columbus, GA.*

For a positive integer n, let a_n and b_n be the unique integers such that

$$(5+\sqrt{3})^n = a_n + b_n\sqrt{3}.$$

Find $\gcd(a_n, b_n)$ as a function of n. Solve the analogous problem when $5+\sqrt{3}$ is replaced by $3+\sqrt{5}$.

2138. *Proposed by Alexandru Girban, Constanta, Romania.*

Let $\triangle ABC$ be a triangle with circumcircle ω and let D be a fixed point on side BC. Let E be a point on ω and let AE meet line BC at F. Find the locus of the circumcenter of $\triangle DEF$ as E varies along ω.

2139. *Proposed by Philippe Fondanaiche, Paris, France.*

Recall that a Pythagorean triple is a triplet of positive integers (a, b, c) such that $a^2 + b^2 = c^2$. We say that a Pythagorean triple is *good* if adding the same single digit to the front of the decimal representations of a, b, and c yields another Pythagorean triple. We will call a Pythagorean triple *very good* if it is good and it is not a nontrivial scalar multiple of another good Pythagorean triple. For example (50, 120, 130) is good, since (150, 1120, 1130) is also a Pythagorean triple, but it is not very good since it is a scalar multiple of the very good triple (5, 12, 13).

Show that there are infinitely many very good Pythagorean triples.

2140. *Proposed by Antonio Garcia, Strasbourg, France.*

For a fixed integer $n \geq 2$, find the minimum value of

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \exp(x_i^2) + \exp\Big(\sum_{1 \leq i < j \leq n} -x_i x_j\Big).$$

Quickies

1117. *Proposed by H. A. ShahAli, Tehran, Iran.*

For every integer $n > 1$, do there exist n integers greater than one such that each integer divides one plus the product of the remaining $n-1$ integers?

1118. *Proposed by George Stoica, Saint John, NB, Canada.*

A number of different objects have been distributed into n boxes $B_1, \ldots, B_n$ with no box empty. All the objects from these boxes are removed and redistributed into $n+1$ new boxes $A_1, \ldots, A_{n+1}$, with no new box empty (so the total number of objects must be at least $n+1$). Prove that there are two objects each of which has the property that it is in a new box that contains fewer objects than its old box.

Solutions

A series involving central binomial coefficients — December 2020

2111. *Proposed by Enrique Treviño, Lake Forest College, Lake Forest, IL.*

Evaluate

$$\sum_{n=0}^{\infty} \frac{\binom{4n}{2n}}{4^{2n}(2n+1)(2n+2)}.$$

Solution by Hongwei Chen, Christopher Newport University, Newport News, VA.
The value of the series is $\frac{4}{3}(\sqrt{2}-1)$. To this end, recall the generating function for the central binomial coefficients

$$\sum_{n=0}^{\infty} \binom{2n}{n} x^n = \frac{1}{\sqrt{1-4x}}, \quad \text{for } |x| < \tfrac{1}{4}.$$

Replacing x by $-x$ gives

$$\sum_{n=0}^{\infty} (-1)^n \binom{2n}{n} x^n = \frac{1}{\sqrt{1+4x}}, \quad \text{for } |x| < \tfrac{1}{4}.$$

Adding these series gives

$$\sum_{n=0}^{\infty} \binom{4n}{2n} x^{2n} = \frac{1}{2}\left(\frac{1}{\sqrt{1-4x}} + \frac{1}{\sqrt{1+4x}}\right).$$

Replacing x by $x/4$ yields

$$\sum_{n=0}^{\infty} \frac{\binom{4n}{2n}}{4^{2n}} x^{2n} = \frac{1}{2}\left(\frac{1}{\sqrt{1-x}} + \frac{1}{\sqrt{1+x}}\right), \quad \text{for } |x| < 1.$$

Integrating this series on $[0, x]$ with $0 < x < 1$, we find

$$\sum_{n=0}^{\infty} \frac{\binom{4n}{2n}}{4^{2n}(2n+1)} x^{2n+1} = \sqrt{1+x} - \sqrt{1-x}.$$

Integrating this series on $[0, x]$ with $0 < x < 1$ again, we find

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{\binom{4n}{2n}}{4^{2n}(2n+1)(2n+2)} x^{2n+2} &= \int_0^x (\sqrt{1+t} - \sqrt{1-t})\mathrm{dt} \\ &= \frac{2}{3}\left((1+x)^{3/2} + (1-x)^{3/2}\right) - \frac{4}{3}.\end{aligned}$$

Applying Abel's convergence theorem and letting $x \to 1$, we conclude

$$\sum_{n=0}^{\infty} \frac{\binom{4n}{2n}}{4^{2n}(2n+1)(2n+2)} = \frac{4}{3}(\sqrt{2}-1)$$

as claimed.

Also solved by Ulrich Abel & Vitaliy Kushnirevych (Germany), Farrukh Ataev (Uzbekistan), Michel Bataille (France), Khristo Boyadzhiev, Paul Bracken, Brian Bradie, Cal Poly Pomona Problem Solving Group, Robert Doucette, Gerald Edgar, Dmitry Fleischman, Mohit Hulse (India), Dixon Jones & Marty Getz, Mark Kaplan & Michael Goldenberg, GWstat Problem Solving Group, Omran Kouba (Syria), Sushanth Sathish Kumar, Elias Lampakis (Greece), Kee-Wai Lau (China), James Magliano, Northwestern University Math Problem Solving Group, Moubinool Omarjee (France), Shing Hin Jimmy Pa (Canada), Angel Plaza (Spain), Rob Pratt, Volkhard Schindler (Germany), Edward Schmeichel, Randy Schwartz, Albert Stadler (Switzerland), Seán M. Stewart (Australia), Ibrahim Suleiman (United Arab Emirates), Michael Vowe (Switzerland), and the proposer. There were two incomplete or incorrect solutions.

A problem from commutative algebra **December 2020**

2112. *Proposed by Souvik Dey, (graduate student), University of Kansas, Lawrence, KS.*

Let R be an integral domain and I and J be two ideals of R such that IJ is a non-zero principal ideal. Prove that I and J are finitely-generated ideals.

Solution by Eugene A. Herman, Grinnell College, Grinnell, IA.
Let $IJ = \langle x \rangle$, where x is a nonzero element of R. Since $x \in IJ$, there exist

$$i_1, \ldots, i_n \in I \text{ and } j_1, \ldots, j_n \in J$$

such that

$$x = i_1 j_1 + \cdots i_n j_n.$$

We claim that

$$I = \langle i_1, \ldots, i_n \rangle \text{ and } J = \langle j_1, \ldots, j_n \rangle.$$

In each of these two equations, it suffices to prove that the left side is contained in the right. For any $i \in I$, there exist $r_1, \ldots, r_n \in R$ such that

$$ij_k = r_k x, k = 1, \ldots, n.$$

Multiply the kth equation by i_k and add the the resulting equations to obtain

$$ix = \sum_{k=1}^{n} r_k i_k x$$

Since R is an integral domain,

$$i = \sum_{k=1}^{n} r_k i_k,$$

and so $I = \langle i_1, \ldots, i_n \rangle$. Similarly, $J = \langle j_1, \ldots, j_n \rangle$.

Also solved by Paul Budney, Noah Garson (Canada), Elias Lampakis (Greece), and the proposer.

A condition for the nilpotency of a matrix **December 2020**

2113. *Proposed by George Stoica, Saint John, NB, Canada.*

Let A be an $n \times n$ complex matrix such that $\det(A^k + I_n) = 1$ for $k = 1, 2, \ldots, 2^n - 1$.

(a) Prove that $A^n = O_n$.

(b) Show that the result does not hold if $2^n - 1$ is replaced by any smaller positive integer.

Solution by Michael Reid, University of Central Florida, Orlando, FL.
(a) First we have a lemma.

Lemma. *Suppose $z_1, \ldots, z_m \in \mathbb{C}$ are such that the power sums $S_k = z_1^k + \cdots + z_m^k$ vanish for $k = 1, 2, \ldots, m$. Then each $z_j = 0$.*

Proof. For $k = 1, 2, \ldots, m$, let σ_k denote the kth elementary symmetric function of $z_1, \ldots, z_m$. By Newton's identities,

$$k\sigma_k = (-1)^{k-1} S_k + \sum_{i=1}^{k-1} (-1)^{i-1} \sigma_{k-i} S_i = 0,$$

for $k = 1, 2, \ldots, m$. Hence each $\sigma_k = 0$. Therefore,

$$(T + z_1)(T + z_2) \cdots (T + z_m) = T^m + \sigma_1 T^{m-1} + \cdots + \sigma_{m-1} T + \sigma_m = T^m.$$

By unique factorization of polynomials, each factor, $T + z_j$, on the left is a constant multiple of T, so each $z_j = 0$. □

The matrix A is similar to an upper triangular matrix M (for example, take M to be a Jordan canonical form of A). Let $d_1, \ldots, d_n$ be the diagonal entries of M. Then $A^k + I_n$ is similar to $M^k + I_n$, which is an upper triangular matrix with diagonal entries $d_1^k + 1, \ldots, d_n^k + 1$, so

$$\det(A^k + I_n) = (d_1^k + 1) \cdots (d_n^k + 1).$$

For each subset $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$, let $b_{\mathcal{S}} = \prod_{j \in \mathcal{S}} d_j$. Thus

$$1 = \det(A^k + I_n) = \prod_{j=1}^{n} (d_j^k + 1) = \sum_{\mathcal{S} \subseteq \{1,2,\ldots,n\}} b_{\mathcal{S}}^k.$$

Let $m = 2^n - 1$, and let $\mathcal{S}_1, \ldots, \mathcal{S}_m$ be the non-empty subsets of $\{1, 2, \ldots, n\}$, and put $z_j = b_{\mathcal{S}_j}$. Then, the equation above becomes $z_1^k + \cdots + z_m^k = 0$, which holds for $k = 1, 2, \ldots, m$. From the lemma, each $z_j = 0$. In particular, for a singleton subset $\{i\}$, we have $d_i = b_{\{i\}} = 0$. Hence M is upper triangular, with all zeros on its diagonal, so its characteristic polynomial is T^n. Since A is similar to M, it has the same characteristic polynomial, so by the Cayley–Hamilton theorem, $A^n = 0_n$.

(b) Let $m = 2^n - 1$, and let $\zeta \in \mathbb{C}$ be a primitive mth root of 1. Let A be the diagonal matrix with diagonal entries $\zeta, \zeta^2, \zeta^4, \ldots, \zeta^{2^{n-1}}$. Then, A is non-singular, so it is not nilpotent. For $k \in \mathbb{N}$, $A^k + I_n$ is the diagonal matrix whose diagonal is

$$\zeta^k + 1, \zeta^{2k} + 1, \zeta^{4k} + 1, \ldots, \zeta^{2^{n-1}k} + 1.$$

Thus,

$$\det(A^k + I_n) = (\zeta^k + 1)(\zeta^{2k} + 1) \cdots (\zeta^{2^{n-1}k} + 1).$$

If k is not divisible by m, then this product telescopes to give

$$\det(A^k + I_n) = \prod_{j=0}^{n-1} \frac{\zeta^{2^{j+1}k} - 1}{\zeta^{2^j k} - 1} = \frac{\zeta^{2^n k} - 1}{\zeta^k - 1} = 1,$$

because $\zeta^{2^n k} = \zeta^{mk+k} = \zeta^k$. Hence,

$$\det(A^k + I_n) = 1$$

for $k = 1, 2, \ldots, 2^n - 2$.

Also solved by Lixing Han & Xinjia Tang, Koopa Tak Lan Koo (Hong Kong), Elias Lampakis (Greece), Albert Stadler (Switzerland), and the proposer. There were two incomplete or incorrect solutions.

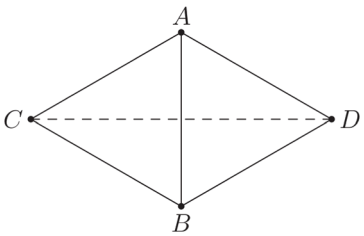
Planar 2-distance sets having four points **December 2020**

2114. *Proposed by Robert Haas, Cleveland Heights, OH.*

Find all configurations of four points in the plane (up to similarity) such that the set of distances between the points consists of exactly two lengths.

Solution by Robert L. Doucette, McNeese State University, Lake Charles, LA.
Suppose A, B, C, D are distinct points in the plane such that the list of six segment distances, AB, AC, AD, BC, BD, and CD, has exactly two real values. For convenience, we may suppose that one of these values is 1. We consider three cases.

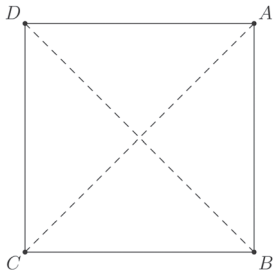
Case 1. Exactly five of the six distances equal 1. Suppose $AB = AC = AD = BC = BD = 1$, $CD \neq 1$. In this case, ABC and ABD must form equilateral triangles. Since $C \neq D$, $ADBC$ must form a rhombus with side length 1 and one pair of opposite angles measuring 60°. This yields a configuration in which the distance not equal to 1 is $CD = \sqrt{3}$.



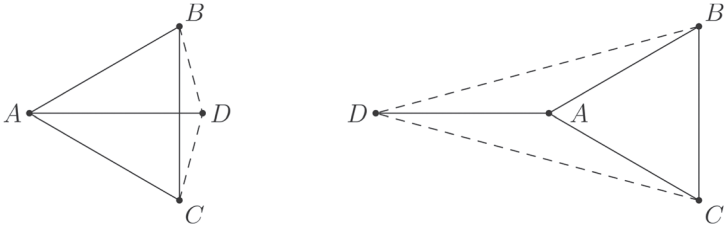
Case 2. Exactly four of the six distances equal 1. There are two subcases to consider.

(i) Suppose first that the two segments with length not equal to 1 do not have an endpoint in common. Say $AC = BD \neq 1$. Since $ABCD$ is a rhombus with congruent diagonals, it must be a square. This yields a configuration in which the distances not equal to 1 are $AC = BD = \sqrt{2}$.

(ii) Suppose next that the two segments with length not equal to 1 do share an endpoint. Say $BD = CD \neq 1$. In this case, ABC forms an equilateral triangle of side length 1. The point D must lie on the perpendicular bisector of segment BC. Either D lies on the same side of $\overleftrightarrow{BC}$ as A or on the opposite side. In the former case, $\triangle BCD$ is a 30°-75°-75° triangle, A is its circumcenter, and $BD = CD = \sqrt{2 + \sqrt{3}}$.

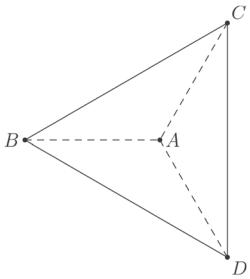


In the latter case, $ABDC$ is a kite with opposite angles of measure 60° and 150°, and $BD = CD = \sqrt{2 - \sqrt{3}}$.

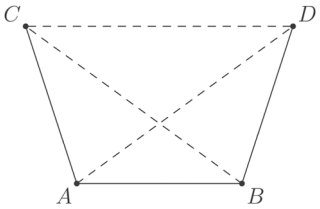


Case 3. Exactly three of the six distances equal 1. Again we consider two subcases.

(i) Suppose that three of the segments of equal length have an endpoint in common. We may assume that $AB = AC = AD = 1$ and $BC = BD = CD \neq 1$. In this case, the points B, C and D lie on the circle with center A and radius 1 and form an equilateral triangle. In other words, BCD forms an equilateral triangle with circumcenter A. In this case, $BC = BD = CD = \sqrt{3}$.



(ii) Next suppose that no three of the segments of equal length share a common endpoint. We may assume that $AB = AC = BD = 1$ and $AD = BC = CD = x > 1$. Since $\triangle ABC \cong \triangle BAD$, $\angle BAC \cong \angle ABD$. If C and D are on opposite sides of $\overleftrightarrow{AB}$, then $ACBD$ is a parallelogram. But by the parallelogram law, $AB^2 + CD^2 = 2AC^2 + 2AD^2$, implying that $1 + x^2 = 2 + 2x^2$, which is impossible. Therefore C and D lie on the same side of $\overleftrightarrow{AB}$ and $ABDC$ is an isosceles trapezoid. Let $m(\angle ADC) = \alpha$. Then $m(\angle BCD) = \alpha$ (since $\triangle ADC \cong \triangle BCD$), $m(\angle ABC) = m(\angle BAD) = \alpha$ (alternating interior angles), $m(\angle ACB) = m(\angle ADB) = \alpha$ (base angles of isosceles triangles), and $m(\angle CAD) = m(\angle CBD) = 2\alpha$ (base angles of isosceles triangles). The sum of the measure of the interior angles of a quadrilateral is 360°, so $10\alpha = 360°$ and $\alpha = 36°$. This means that A, B, C, and D are four of the five vertices of a regular pentagon. In this case, $AD = BC = CD = (1 + \sqrt{5})/2$.



We have shown that there are six configurations of four points satisfying the requirements described in the problem statement: (1) a rhombus with one pair of opposite angles measuring 60°, (2) a square, (3) an isosceles triangle with vertex angle of 30° and its circumcenter, (4) a kite with a pair of opposite angles measuring 60° and 150°, (5) an equilateral triangle and its circumcenter, and (6) four of the five vertices of a regular pentagon.

Also solved by Diya Bhatt & Riley Platz & Tony Luo (students), Viera Cernanova (Slovakia), M. V. Channakeshava (India), Seungheon Lee (Korea), Eagle Problem Solvers, Michael Reid, Celia Schacht, Albert Stadler (Switzerland), Tianyue Ruby Sun (student), Randy K. Schwartz, and the proposer. There were six incomplete or incorrect solutions.

Two compass and straightedge constructions **December 2020**

2115. *Proposed by H. A. ShahAli, Tehran, Iran.*

Let A and B be two distinct points on a circle and let k be a positive rational number.

(a) Give a compass and straightedge construction of a point C on the circle such that $AC/BC = k$.

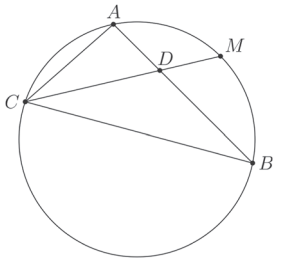
(b) Give a compass and straightedge construction of a point C on the circle such that $AC \cdot BC = k$. As part of your solution, find the restrictions on k in terms of AB and the radius of the circle necessary for such a C to exist.

Solution by Enrique Treviño, Lake Forest College, Lake Forest, IL.

(a) It is well known that we can construct a point D on segment AB such that $AD/BD = k$. Let M be a point of intersection of the perpendicular bisector of AB with the given circle. Then $AM = BM$. Let C be the second point of intersection of $\overleftrightarrow{MD}$ with the circle. Since $AM = BM$, then $\angle ACM = \angle BCM$. Therefore D is on the angle bisector of $\angle ACB$ and by the angle bisector theorem

$$\frac{AC}{BC} = \frac{AD}{BD} = k.$$

An alternative solution is to note that $\{X | AX/BX = k\}$ is a circle or the perpen-



dicular bisector of AB. This curve is readily constructible, and we then find its intersection with the original circle.

(b) To solve this problem in full generality, we need a segment of length 1 to be given. It is well known that given such a segment, $\lambda \in \mathbb{Q}^+$, and a segment of length a, we can use similar triangles to construct a segment of length λ/a.

Denote the center of the circle by O and let $m\angle AOB = 2\alpha$. Then $x = 2r\sin\alpha$. For any point C on the circle, $m\angle ACB = \alpha$ or $m\angle ACB = \pi - \alpha$. In either case $\sin\angle ACB = \sin\alpha$. We will denote the area of $\triangle PQR$ by (PQR). We know

$$(ABC) = \frac{AC \cdot BC \cdot \sin\alpha}{2}.$$

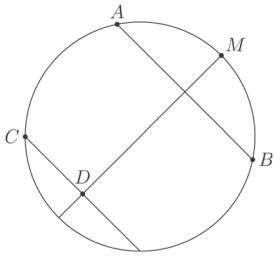
Let h be the height of $\triangle ABC$ with base AB, then

$$(ABC) = \frac{xh}{2}.$$

Therefore

$$AC \cdot BC = 2rh.$$

Let ℓ be the perpendicular bisector of AB. Using the facts stated above, we can construct a point D on ℓ such that the distance from D to AB is $k/(2r)$. Next, we draw a line through D parallel to $\overleftrightarrow{AB}$ and let C be one of the points of intersection of this line with the given circle. This point C satisfies $AC \cdot BC = 2rh = k$.



The largest possible value of $AC \cdot BC$ occurs when C lies on the perpendicular bisector of AB at a maximum distance from AB, namely when

$$AC = BC = r + \sqrt{r^2 - \frac{x^2}{4}}.$$

Therefore

$$\begin{aligned} k &= AC \cdot BC \\ &\le \left(r + \sqrt{r^2 - \frac{x^2}{4}}\right)^2 \\ &= r\left(2r + \sqrt{4r^2 - x^2}\right). \end{aligned}$$

Also solved by Michel Bataille (France), Ivko Dimitrić, Elias Lampakis (Greece), Celia Schacht, Albert Stadler (Switzerland), and the proposer. There were three incomplete or incorrect solutions.

Answers

Solutions to the Quickies from page 74.

A1117. The answer is yes.
We claim that the integers $a_1 = 2$ and $a_k = a_1 a_2 \cdots a_{k-1} + 1$ for all $k = 2, \ldots, n$ satisfy the conditions of the problem.

Since a_i is odd, for $i > 1$,

$$a_1 | a_2 \cdots a_n + 1,$$

and clearly

$$a_n | a_1 \cdots a_{n-1} + 1.$$

Now, if $n > 2$, then for any $k = 2, \ldots, n-1$, we have

$$a_{k+1} \equiv 1 \pmod{a_k}, \ldots, a_n \equiv 1 \pmod{a_k}.$$

Therefore

$$\begin{aligned} a_1 \cdots a_{k-1} \cdot a_{k+1} \cdots a_n + 1 &\equiv a_1 \cdots a_{k-1} \cdot 1 \cdots 1 + 1 \\ &\equiv a_k \\ &\equiv 0 \pmod{a_k}, \end{aligned}$$

as claimed.

Note: The search for other solutions to this problem is known as the improper Znám problem.

A1118. Represent the situation as a bipartite graph where the vertices are $\{B_i\} \cup \{A_j\}$ and an edge connects B_i to A_j if there is an object that is moved from B_i to A_j.

Fix n and induct on k, the total number of edges. Note that $k \geq n + 1$.

Base case: If $k = n + 1$ exactly one of the B_i's has degree two; then the rest have degree one. That B_i must be adjacent to an A_k and an A_ℓ each with degree one. Those edges represent the desired objects.

Induction step: Suppose the result holds for $k - 1$ edges. Consider the situation with k edges. If the graph has two adjacent vertices, each with degree greater than one, then remove the edge connecting them. The new graph satisfies the induction hypothesis and the result follows.

If not, then every edge is adjacent to a vertex with degree one. Say there are s degree one B_i's adjacent to a degree one A_j, t degree one B_i's adjacent to an A_j not of degree one, and t B_i's not of degree one that are adjacent to degree one A_j's. Then $s + t + u = n$. If $u = 0$, then $s + t = n$, and the total number of edges is $k = s + t$. Since $k \geq n + 1$ this is impossible. Therefore, $u > 0$ and we are done, since we can pick one of the B_i's in this class and two of the edges connected to that B_i to represent the desired objects.

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.

Ellenberg, Jordan, *Shape: The Hidden Geometry of Information, Biology, Strategy, Democracy, and Everything Else*, Penguin Press, 2021; 463 pp, $28. ISBN 978-1-9848-7905-9.

"We are living in a wild geometric boomtown, global in scope. ... There's something special about geometry, something that makes it worth writing poems about." Or, in author Ellenberg's case, writing a book. It is filled with essays on various aspects of geometry (Euclid, topology, symmetries, random walks, trees, strategy spaces, maps, geometric progressions, the golden ratio, redistricting), all richly embedded in historical narratives accompanied by delicious side details (e.g., I didn't know that Andrei Markov demanded to be excommunicated by the Russian Orthodox Church—and was). My favorite chapter was "How many holes does a straw have?"—if you think you know how many, this chapter may be a revelation, as it was for me. This book is an absolutely splendid, easily understandable, and utterly engaging exposition of mathematical ideas.

Sumpter, David, *The Ten Equations That Rule the World and How You Can Use Them Too*, Flatiron Books, 2021; 247 pp, $28.99. ISBN 978-1-250-24696-7.

I would have thought that the equation of gravitational attraction "rules the world"; but you won't find it here amid "The Judgment Equation," "The Influencer Equation," "The Reward Equation," and seven others. Author Sumpter presents these equations as offering "the seeds of success, popularity, wealth, self-confidence, and sound judgment," and he proposes them as having been mastered and passed down by a "secret society" through the ages. This quasi-mystical approach is a bit of a put-on, but it may hook some readers into reading the book. Sumpter is a professor of applied mathematics and author of *Soccermatics* (on soccer and gambling on it) and *Outnumbered* (on the influence of algorithms on our lives). The actual equations include Bayes's formula, logistic regression, confidence intervals, the Markov assumption, and a stochastic differential equation. Sumpter weaves examples of their use into authentic and interesting narratives from his experiences. The book ends, however, on a counterpoint, with an ethical discussion of utilitarianism, logical positivism, and the pre-eminence of moral intuition, plus the all-important caution that "there are no purely mathematical answers to questions about real life."

Riehl, Emily, Infinite math, *Scientific American* 325 (4) (October 2021) 32–41. https://www.scientificamerican.com/article/infinity-category-theory-offers-a-birds-eye-view-of-mathematics1/.

"If mathematics is the science of analogy, the study of patterns, then category theory is the study of mathematical patterns of thought." Author Riehl asserts that abstraction—taking a broader view—is the key to how "mathematicians can quickly teach each new generation of undergraduates discoveries that astonished the previous generation's experts." In particular, the language of category theory provides common proofs of theorems in varying branches of mathematics. Riehl nimbly explains the concepts of isomorphism, homotopy, category, and the generalization to an ∞-category.

Math. Mag. **95** (2022) 83–84. doi:10.1080/0025570X.2022.2001230

Enright, Sam, A layman's guide to recreational mathematics videos, https://samenright.com/2021/08/31/a-laymans-guide-to-recreational-mathematics-videos/.

There are many inspiring books about mathematics for the lay person, such as the one by Jordan Ellenberg reviewed above. But the taste of the younger generation (and maybe of some older folks, too) is oriented toward video presentations. The internet is chock full of videos, but what's worth watching if you want to learn about mathematics? Author Enright ignores formal education (tutorials) and high-quality lectures and divides the remaining universe into explainer channels (including Numberphile), problem-solving channels (you can watch a recorded livestream of solving integrals for six hours straight), other channels that sometimes do math (finding the eigenvalues of a Möbius strip), music and fun (Tom Lehrer, sure; but a musical proof of why e is irrational?!), and podcasts (e.g., "The Joy of X" by Steven Strogatz).

Carlisle, Justus, Kyle Hammer, Robert Hingtgen, and Gabriel Martins, Skateboard tricks and topological flips, https://arxiv.org/abs/2108.06307.

In watching the Olympic skateboard finals, I did not observe great variety in the tricks performed. The authors model skateboard flip tricks as continuous curves in the group SO(3) of rotations about the origin of three-dimensional Euclidean space. They conclude that up to continuous deformation, there are only four flip tricks.

Kubota, Taylor, Stanford algorithm helps modern quilters focus on creativity, https://news.stanford.edu/2021/06/02/new-algorithm-modern-quilting/.

Leake, Mackenzie, Gilbert Bernstein, Abe Davis, and Maneesh Agrawala, A mathematical foundation for foundation paper pieceable quilts, *ACM Transactions on Graphics* 40 (4) Article 65, August 2021. https://web.stanford.edu/~mleake/projects/paperpiecing/files/FPP_small.pdf.

I have admired quilts, but until this article I did not appreciate the art involved in creating them—in particular, the art of sewing the seams of the pieces so as not to show on the upper surface. Quilters begin with a geometric design and an order in which the pieces should be sewn into the quilt. Sometimes the order can be formulated into a pattern guide printed on paper that is used as a foundation in sewing the pieces. When can a design be developed into such a foundation paper piece? When the dual hypergraph of the design is acyclic, as the authors prove. For such designs, they offer an algorithm that can generate sewing orders, a corresponding interactive design tool, and examples.

Hartnett, Kevin, New math book rescues landmark topology proof, https://www.quantamagazine.org/new-math-book-rescues-landmark-topology-proof-20210909/.

Should a mathematical result be regarded as true if its author just convinces a few experts that it is true but never writes out a proof that is understandable more widely? The Poincaré conjecture in n dimensions is that every simply-connected closed n-manifold is homeomorphic to the n-sphere. In 1981 Michael Freedman, after seven years of solitary work, announced a proof of it for $n = 4$. (It had been settled in 1961 by Stephen Smale for $n > 4$, and the case $n = 3$ was later resolved by Grigori Perelman in 2003.) Freedman convinced colleagues and provided an outline but never wrote out a proof "that people who had never met him could read and learn on their own." At last, a team of mathematicians has done that, in *The Disc Embedding Theorem*, edited by Stefan Behrens et al. (Oxford University Press, 2021).

Hartnett, Kevin, The mystery at the heart of physics that only math can solve, https://www.quantamagazine.org/the-mystery-at-the-heart-of-physics-that-only-math-can-solve-20210610/.

The "standard model" of quantum field theory, combining 12 fundamental particle fields and four force fields into a single equation, has been enormously successful in explaining experiments. But what is a quantum field theory? What should the basic objects be? What should be their properties? How should they interact? The mathematics must involve the interactions of infinite-dimensional operators. Physicists are eager for mathematicians to develop a complete mathematical description, but it may take generations.