GATHERING 4 GARDNER 13

G4G13 Exchange Book

Art, Games, Magic, & Math

Atlanta, Georgia

APRIL 11 - APRIL 15, 2018

VOLUME 1

G4G13 Exchange Book

VOLUME 1

The Gift Exchange is an integral part of the Gathering 4 Gardner biennial conferences. Gathering participants exchange gifts, papers, puzzles and other interesting artifacts. This book contains gift exchange papers from the conference held in Atlanta, Georgia from Wednesday, April 11th through Sunday, April 15th, 2018. It combines all of the papers offered as exchange gifts in two volumes.

Acknowledgments

Gathering 4 Gardner would like to offer thanks to the following individuals:

Freddy Bendekgey – for editing and laying out the pages of this book
The Creative Fold, LLC – for project management and art direction
The G4G Board of Directors – for making this publication possible
Nancy Blachman
Erik Demaine
Colm Mulcahy
Robert Crease
James Gardner
Jade Vinson
Vickie Kearn

There are many things you can buy or make yourself, but there are some things that can only be created by a group. This book is such an item.

In Memoriam

With sadness, we note the passing of the following Gathering 4 Gardner attendees:

Robert Abbott
Elwyn Berlekamp
Markus Götz
Marc Pelletier

Unless otherwise noted, all papers appear unedited, exactly as they were submitted.

Preface

by Doris Schattschneider

Welcome to this glimpse of some of the fun and excitement of the 13th Gathering for Gardner (G4G13) in Atlanta, Georgia, April 11-15, 2018. Here you will find the program of events, and 78 papers that are write-ups by many of the presenters who made this event so vibrant. The subjects are far-ranging, all touching on subjects that fascinated Martin Gardner. Placed into sections on Art, Games, Math, and Magic (difficult decisions for the organizers, since many fall into more than one, or possibly none of the categories), these papers describe puzzles, games, illusions, magic, and curiosities both mathematical and otherwise.

Several presentations at G4G13 were not written up for these two volumes, but videos of most of those can be found at the web site: **https://www.gathering4gardner.org/g4g13-videos/**.

Here you will find the conference talks by Doris Schattschneider (Marjorie Rice's fruitful search for tiling pentagons,) Ernő Rubik (tales of his famous cube), and Manjul Bhargava (magic with magical numbers); moving memorial tributes to Solomon Golomb, Raymond Smullyan, and Marc Pelletier; reminisces by Jim Gardner of his father Martin; interviews with many of the conference participants; and much more.

Beyond the written papers and videos, many wonderful happenings were recorded only in the minds of participants: the amazing feats of magicians Max Maven and Carlos Vinuesa, followed by close-up magic tables, and the foot-juggling with almost impossible body twisting by circus artist Roxana Küwen. Perhaps most of all, the best "takeaways" were the unexpected encounters of like minds sharing excitement over a new discovery and the thorough enjoyment of an intellectual circus.

This year G4G activities expanded beyond the venue of the Ritz Carlton with several public events. On the eve of the conference opening, Ernő Rubik held forth and answered questions at Georgia Tech's Clough Auditorium. On the second day, on a beautiful sunny afternoon, participants picnicked in the neighboring town of Decatur, where kids (and kids at heart) could help build amazing large sculptures and make and take all kinds of fun items. Across the street, Sarah Garvin Rodgers welcomed everyone at her gallery, Different Trains, for a wonderful exhibition of original prints by M.C. Escher. This was complemented by Doris Schattschneider's evening lecture, "The Mathematics of M.C. Escher's Art," at Decatur Presbyterian Church. And on the conference's last day, all were treated to Manjul Bhargava's performance on "Poetry, Drumming, and Mathematics" at Fulton County Central Library.

Martin Gardner sparked a myriad of followers with his writings on unpredictable subjects, and this conference in his honor brought together amateurs and professionals, old and young (one presenter was 8 years old), having one common trait—an insatiable curiosity and desire to share.

Doris Schattschneider

G4G13 video presentations can be found on our website: **www.gathering4gardner.org/g4g13-videos/**

Table of Contents | VOLUME 1

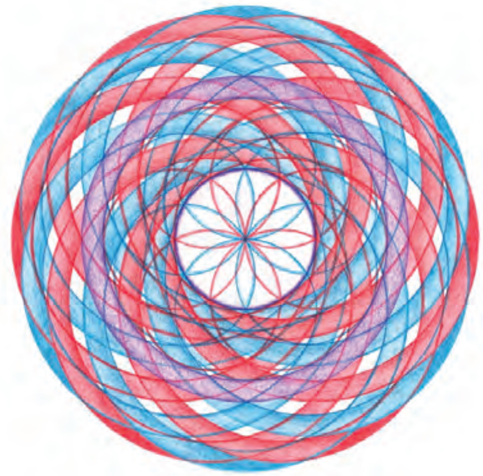
Table of Contents (cont.)

MAGIC

MATH

Unofficial Logos

Submitted by Robert Fathauer



Submitted by Vandorn Hinnant

Submitted by Teja Krasek

Submitted by Miguel Palomo

Presentation Schedule

Thursday, April 12th, 2018

Morning Session: 8:30 AM - 12:00 PM

FEATURED PRESENTATION

	Speaker	Title		
8:30 AM	Brian Hopkins	Difference Dice	*6min*	8:30 AM
	Ryuhei Uehara	Design Schemes for Fair Dice	*6min*	
	James Grime	Nontransitive dice for three players	*6min*	
	Kenneth Brecher	Introducing the PiTOP	*6min*	
	Henry Segerman and Sabetta Matsumoto	Design of hinged 3D auxetic mechanisms	*6min*	
	David Nacin	Solutions to Klein Four Puzzles	*6min*	
	Adam Rubin	Functional Impossible Objects	*6min*	
	Jane Kostick	13 Piece Puzzles	*6min*	
	Bernardo Recaman	Ramanujan Sums	*6min*	
	Akio Hizume	Fibonacci Jigsaw Puzzle etc.	*6min*	
	Hirokazu Iwasawa	Classic False Coin Puzzles without Mathematical Induction	*6min*	
		Break	*30min*	
	Doris Schattschneider	The Story of Marjorie Rice	*45min*	
	Jaap Scherphuis	Developments in Pentagonal tilings	*5min*	
	Bob Hearn	There are only 15	*6min*	
	Laura Taalman	Printing Perfect Pentagons	*6min*	
	Kate McKinnon	Engineering and Art: The Magic of the Kaleidocycle and the Contemporary Geometric Beadwork project	*6min*	
	Jordan Gold and Elan Lee	How we built the world's smartest computer with a few pieces of cardboard and a pineapple	*10min*	
12:00 PM				12:00 PM

Afternoon Session: 1:30 PM - 5:30 PM

	Speaker	Title		
1:30 PM	John Rausch, Nick Baxter, and Bill Cutler	A Fireside Chat with Stewart T. Coffin	*27min*	1:30 PM
	David Michael Greene	Self-Construction in Conway's Life	*6min*	
	Adam P. Goucher	Evolving lifeforms on lattices	*6min*	
	Tomas Rokicki	HashLife	*6min*	
	James Emmett Gardner	Growing up Around Martin Gardner: Another Round	*30min*	
	Cindy Lawrence	Play Truchet: Using the Truchet tiling to engage the public with mathematics	*6min*	
	R. William Gosper	True planefills versus "Spacefilling Curves"	*6min*	
	Gary Antonick	Thirteen Bounces	*5min*	
	Yossi Elran	G4G's Celebration of Mind -- exciting the public and expanding MG's legacy	*4min*	
	Louis Hirsch Kauffman	Rope Tricks and Topology	*6min*	
		Break	*30min*	
	Eleftherios Pavlides	The Geometry of Motion of the Chiral Icosahedral Hinge Elastegrity	*6min*	
	Robert Bosch	Figurative Subgraphs	*6min*	
	Debora A Coombs	Geometry in five-dimensions: Building quasicrystals from Penrose tiling	*6min*	
	Gwen Fisher	Beaded Tesselations of Polyhedra and Penrose Tilings	*6min*	
	Jim Propp	You Can't Count to Thirteen ...	*6min*	
	Spandan Bandyopadhyay	Hexprimes	*6min*	
	Erik Demaine	Sliding Coins	*6min*	
	Gabriel Doran Kanarek	Repetitive Patterns in the Juggler Sequence	*6min*	
	Alex Bellos	Puzzle Ninja	*5min*	
	William Gasarch	The Muffin Problem	*5min*	
	Stephen Wolfram	TBA	*6min*	
5:30 PM				5:30 PM

Presentation Abstracts Available Online: www.gathering4gardner.org/g4g13-abstracts.pdf

Presentation Schedule

Friday, April 13th, 2018

FEATURED PRESENTATION

Morning Session: 8:30 AM - 11:15 AM

8:30 AM

Speaker	Title	
Miguel Palomo	The Sudoku Ripeto Family	*5min*
James Lee	A Slick Solution to the Tax Man Problem	*6min*
Elwyn Berlekamp	Sol Golomb tribute video	*6min*
Donald Bell	Loyd Polyominoes	*6min*
Aaron Siegel	Polyformer	*6min*
Tom van der Zanden	Packing polyominoes into a 3-by-n box is as hard as it gets	*6min*
Anany Levitin	Polyomino Puzzles and Algorithm Design Techniques	*5min*
Stewart Temple Coffin	Martin's Menace	*4min*
Robert P Crease	Super Thirteen: Welcome to G4G's Teenage Years!	*5min*
Colin Wright	Thirteen - Insufficiently Maligned	*5min*
	Break	*30min*
Erno Rubik	Cubic Tales	*45min*
George Hart	Mathematical Construction Activities During the Excursion	*10min*

11:15 AM

Presentation Abstracts Available Online: www.gathering4gardner.org/g4g13-abstracts.pdf

Presentation Schedule

Saturday, April 14th, 2018

FEATURED PRESENTATION

Morning Session: 8:30 AM - 12:00 PM

	Speaker	Title		
8:30 AM	Akira Nishihara	Geometric toys	*6min*	8:30 AM
	Kate Jones	Lucky 13	*6min*	
	Scott Kim	Motley Dissections	*6min*	
	Rochelle Kronzek Miller	Raymond and Martin -- Pioneers and Legends	*30min*	
	Kokichi Sugihara	Evolution of Impossible Objects	*6min*	
	Stephen Macknik	Champions of Illusion	*6min*	
	Alexa Meade	TBA	*5min*	
	Sabetta Matsumoto and Henry Segerman	Non-euclidean virtual reality	*12min*	
		Break	*30min*	
	Manjul Bhargava	Some Magical Numbers and Their Uses in Magic and Mathematics	*45min*	
	Robert P Crease	How is Science Denial Possible?	*6min*	
	Mike Naylor	The Last Crumb	*6min*	
	Derrick Chung	An Elegant Solution to Rusduck's "A Study in Stud"	*6min*	
	Vladimir Bulatov	Cross sections of three dimensional hyperbolic tilings	*6min*	
	Elwyn Berlekamp	Gallimaufry of Games	*6min*	
12:00 PM				12:00 PM

Afternoon Session: 1:30 PM - 5:30 PM

	Speaker	Title		
1:30 PM	Dana S Richards	Martin Gardner, Annotator	*30min*	1:30 PM
	Thomas Francis Banchoff	Foxtrot Half-Empty Half-Full Problem, including 13	*6min*	
	Theodore Gray	Mechanical Gifs	*6min*	
	Dana Mackenzie	2184 (Oh, the Absurdity)	*6min*	
	Paul Hidebrandt and Amina Buhler-Allen	A tribute to Marc Pelletier	*20min*	
	George I. Bell	The Sailing Stones of Death Valley	*6min*	
	Howard I Cannon	Demonstrating math with carefully drawn triangles in JavaScript	*6min*	
	Raymond Hall	Physics Fun: The use of Social Media as a Museum of Science and Math	*6min*	
	Margaret Kepner	4 x 13	*6min*	
	Duane A Bailey	A Grammatical Approach to the Curling Number Conjecture	*6min*	
		Break	*30min*	
	Adam Atkinson	Medieval French Poetry	*5min*	
	Tanya Khovanova	Crypto and Fractal Word Searches	*6min*	
	Colm Mulcahy	Martin Gardner Word Play	*5min*	
	Joshua Pines	Scrabble Seven-letter Words	*6min*	
	Tom Bessoir	Prime Perfect	*6min*	
	Rik van Grol	Balance puzzles -- you either love them or curse them	*5min*	
	Stuart Moskowitz	Lewis Carroll Should Have Taught Sixth Grade Math	*6min*	
	Pete McCabe	Persistimis Possessiamo	*6min*	
	Robert Fathauer	Knotting and Numbering Kite Tiling Rosettes	*6min*	
	Robert W Vallin	Maverick Solitaire and Three-Card Poker	*6min*	
	Lyman Porter Hurd	Kadon's Dezign-8	*6min*	
	Nancy Blachman	How I Finagled nearly 13 Invitations to White House Events in 2016	*5min*	
5:30 PM				5:30 PM

Presentation Abstracts Available Online: www.gathering4gardner.org/g4g13-abstracts.pdf

Presentation Schedule

Sunday, April 15th, 2018

Morning Session: 8:30 AM - 11:30 AM

	Speaker	Title		
8:30 AM	Dick Esterle	ALMOST 13 - ICOSAHEDRON let me count the ways	*6min*	8:30 AM
	Gerard Westendorp	Hinged polyhedra and hinged tessellations.	*6min*	
	Yossi Elran	13 ways to tie a knot in a strip of paper	*5min*	
	Peter Knoppers	The elusive 13 piece complete set puzzle	*5min*	
	Michael Tanoff	Mr. Apollinax's Wedding Ring	*6min*	
	Susan Goldstine	Symmetry Samplers	*6min*	
	Gergo Kiss	The Derivation of the Spidron Formula	*6min*	
	Ricardo Teixeira	Si Stebbins and Group Theory	*6min*	
	Ann Schwartz	Making Waves: The Pentaflexagon	*6min*	
	Andrew John Rhoda	The Slocum Mechanical Puzzle Collection at the Lilly Library	*6min*	
	Colin Beveridge	An Old-Timey Cipher	*6min*	
	Karl Schaffer	Edgy Puzzles	*6min*	
	Hilarie Orman	Mental Factoring	*6min*	
	Carolyn Yackel	Enough Lace Patterns for Fibonacci	*6min*	
	Henry Strickland	Rainy Day Stories (in the Temporal Logic called D)	*6min*	
	Oded Margalit	PonderThis	*6min*	
	Ben Chaffin	A Tale of Two Powers: Finding zeros in powers of 2	*6min*	
	ET Trigg	Bidding in Contract Bridge considered as a communication channel	*6min*	
	Greg N. Frederickson	Hidden in Plane Sight: the Extraordinary Vision of Ernest Irving Freese	*6min*	
	Ryan William Morrill	Playing nice with a weighted coin	*6min*	
	Lucas Garron	The Perfect Easing Function	*6min*	
	Ron Taylor	Mathematics of color addition games	*6min*	
	Roger Russell Manderscheid	Misconceptions of Measurement: Time, Space and Numbers	*5min*	
11:30 AM				11:30 AM

Afternoon Session: 12:30 PM - 2:00 PM

	Speaker	Title		
12:30 PM	James Joseph Solberg	(Phi)ve is Magic	*6min*	12:30 PM
	Robert Munafo	Using the Analogue Approximation Finder	*6min*	
	David Hall	Recipe for a 'bola Honeycombs	*6min*	
	George Hart	Knotted Bracelet	*6min*	
	William G Ames	Extending the 10958 problem	*6min*	
	Nathaniel Wing Segal	Deconstructing Magic Squares	*6min*	
	Masayoshi Iwai	Tilings of Equilateral Tridecagons	*6min*	
	Carl N. Hoff	From Untouchable 11 to Hazmat Cargo	*6min*	
	Yoshiyuki Kotani	Tiling of 123456-edged hexagon	*6min*	
	Charles Bernard Sonenshein	How Martin Gardner Helped Me Keep My Sanity as a Math Teacher for 50 Years	*6min*	
	Barney Sperlin	How Safe Is It?	*6min*	
2:00 PM				2:00 PM

Presentation Abstracts Available Online: www.gathering4gardner.org/g4g13-abstracts.pdf

G4G13 ADDENDUM

REVISION

Revisions to the Presentation Schedule for THURSDAY, April 12th:

Adam Rubin	Functional Optical Illusions	6min

	Afternoon Break	
Rochelle Kronzek Miller, Jason Rosenhouse, and Elizabeth Carpenter	Honoring the Late, Great Storyteller of Logic - Raymond Smullyan	30min
Eleftherios Pavlides	The Geometry of Motion of the Chiral Icosahedral Hinge Elastegrity	6min

Revisions to the Presentation Schedule for FRIDAY, April 13th:

Robert P Crease	Super Thirteen: Welcome to G4G's Teenage Phase!	6min

Revisions to the Presentation Schedule for SATURDAY, April 14th:

Scott Kim	Motley Dissections	6min
Robert Bosch	Figurative Subgraphs	6min
Spandan Bandyopadhyay	Hexprimes	6min
Alex Bellos	Puzzle Ninja	5min
Jim Propp	You Can't Count to Thirteen ...	6min
Kokichi Sugihara	Evolution of Impossible Objects	6min
Stephen Macknik	Champions of Illusion	6min
Alexa Meade	2D+3D	5min

Manjul Bhargava	Some Magical Numbers and Their Uses in Magic and Mathematics	45min
Carlos Vinuesa	Dealing with Shuffles	6min
Robert P Crease	How is Science Denial Possible?	6min

Lunch Break

	Speaker	Title		
1:30 PM	Dana S Richards	Martin Gardner, Annotator	18min	1:30 PM
	Darren Glass	Chutes and Ladders Without Chutes or Ladders	6min	
	Peter Winkler	Puzzles that Solve Themselves	6min	
	Thomas Francis Banchoff	Foxtrot Half-Empty Half-Full Problem, including 13	6min	

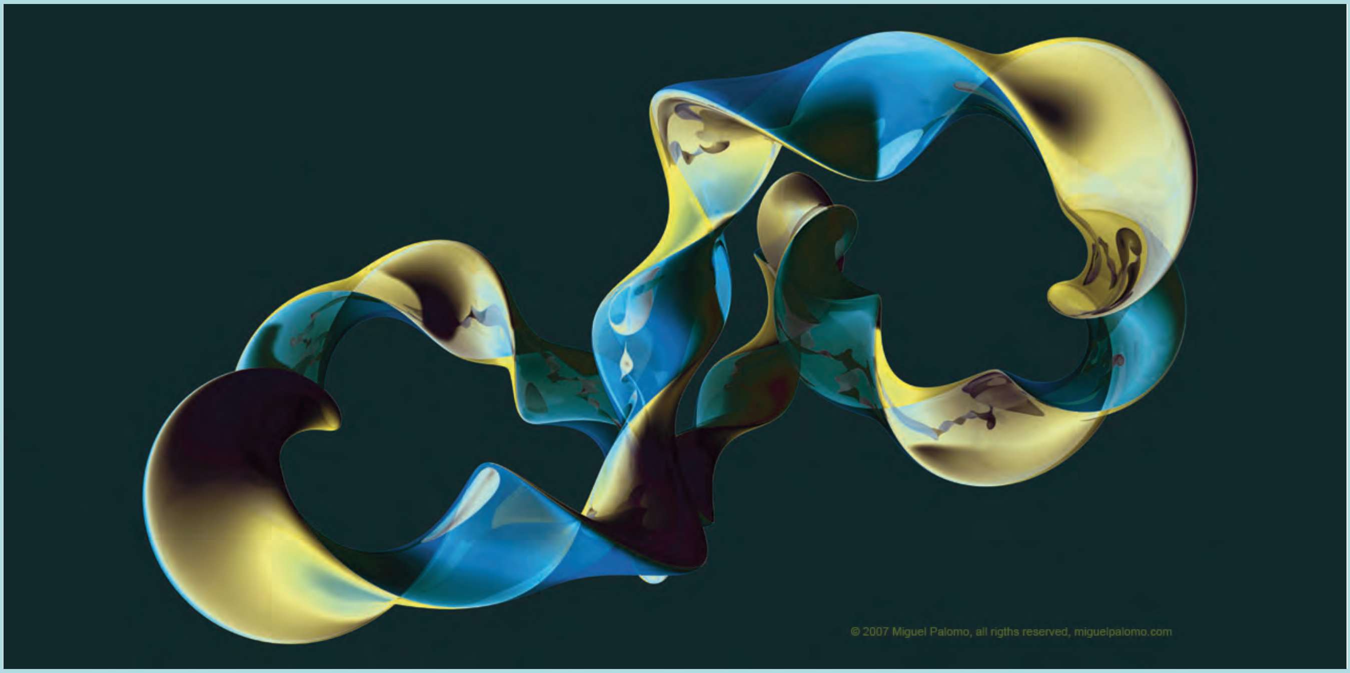
Tom Bessoir	prime perfect	6min
Rik van Grol	Balance puzzles -- you either love them or curse them	5min
Stuart Moskowitz	Lewis Carroll Should Have Taught Sixth Grade Math	6min
James Joseph Solberg	(Phi)ve is Magic	6min
Robert Fathauer	Knotting and Numbering Kite Tiling Rosettes	6min

Revisions to the Presentation Schedule for SUNDAY, April 15th:

Lunch Break

	Speaker	Title		
12:30 PM	Pete McCabe	Persistimis Possessiamo	6min	12:30 PM
	Robert Munafo	Using the Analogue Approximation Finder	6min	
	David Hall	Recipe for a 'bola Honeycombs	6min	
	George Hart	Tying the Knot	6min	
	Bill Ames	Extending the 10958 problem	6min	
	Nathaniel Segal	Deconstructing Magic Squares	6min	

ART



Math Art | Miguel Palomo | Page 59

Card Shuffling Visualizations

Roger Antonsen

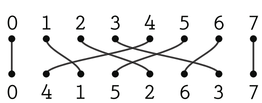
University of Oslo, Norway; rantonse@ifi.uio.no

Abstract

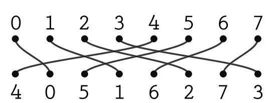
This paper discusses a few natural ways to visualize card shuffles, like the *perfect in- and out-shuffle* and the *milk shuffle*. These visualizations are designed to highlight interesting properties of the shuffles, like the *order* of the shuffles and the *stay-stack principle*. Such properties have attracted magicians and mathematicians alike, and visualizations like these make it possible to see properties that would have been hard to see otherwise.

Four Ways to Shuffle a Deck of Cards

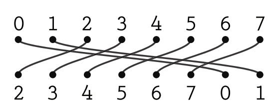
Imagine that you have a deck of cards, say eight cards numbered from 0 to 7: [0, 1, 2, 3, 4, 5, 6, 7]. Split the deck precisely into two halves, [0, 1, 2, 3] and [4, 5, 6, 7], and then interleave the cards evenly such that the top card remains at the top and the bottom card remains at the bottom. The result is [0, 4, 1, 5, 2, 6, 3, 7]. This is called a *perfect out-shuffle*. The corresponding *perfect in-shuffle* is the interleaving of the two halves such that the top card becomes the *second* card from the top: [4, 0, 5, 1, 6, 2, 7, 3]. (Perfect shuffles are also referred to as *faro* shuffles.) Another way to shuffle a deck is to give it a *cut*. This is where the top part of the deck, say [0, 1], is placed at the bottom, giving us [2, 3, 4, 5, 6, 7, 0, 1]. Finally, we have the *milk shuffle*. This is where we repeatedly pinch the deck and pull off the top and the bottom card into a separate pile. This move, often referred to as milking the deck, is repeated until all the cards have been pulled off. In this case, the first two cards to pulled off is [0, 7]. Then, [1, 6], [2, 5], and finally [3, 4], which becomes the two top cards: [3, 4, 2, 5, 1, 6, 0, 7]. A nice and clean way to visualize these shuffles is with the following type of diagrams:



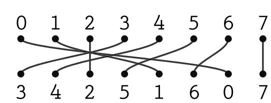
perfect out-shuffle



perfect in-shuffle



two-card cut



milk shuffle

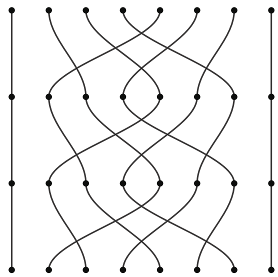
The top rows represent the cards in a deck *before* each shuffle, and the bottom rows represent the order *after* each shuffle. Observe for example that in a perfect out-shuffle, the top cards remain on the top, while in a perfect in-shuffle, the top card becomes the second card. The diagrams may also be read from *below*, or – equivalently – turned upside-down, in which case the shuffles are called the respective *reverse* shuffles. A reverse perfect shuffle occurs naturally when cards are being dealt or separated into distinct piles.

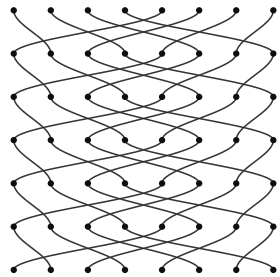
Permutations and the Order of Card Shuffles

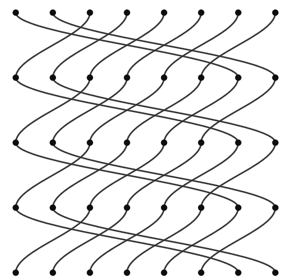
From a mathematical point of view each shuffle is a *permutation* of the cards, and repeated shuffling can be viewed as the *composition* of permutations. Visually, this corresponds to stacking diagrams like this on top of each other. In more mathematical terms, the perfect out-shuffle on a deck of N cards, where N is even, moves the card in position p to the position $\mathbf{O}(p) \equiv 2p \pmod{N-1}$ for all p less than $N-1$ and to $\mathbf{O}(N-1) = N-1$ when p is $N-1$. In a similar way, the perfect in-shuffle moves the card in position p to $\mathbf{I}(p) \equiv 2p+1 \pmod{N+1}$. A cut of c cards simply moves the card in position p to $\mathbf{C}(p) \equiv p-c \pmod{N}$.

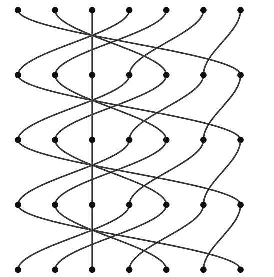
We will now consider what happens as we *repeat* the various shuffles. After a certain number of repetitions, called the *order* of each shuffle, the deck will return to its original order. The order of a particular

shuffle depends on the number of cards: For eight cards, this happens after three times for the out-shuffle, six times for the in-shuffle, four times for the two-card cut, and four times for the milk shuffle:


3 *out-shuffles with* 8 *cards*

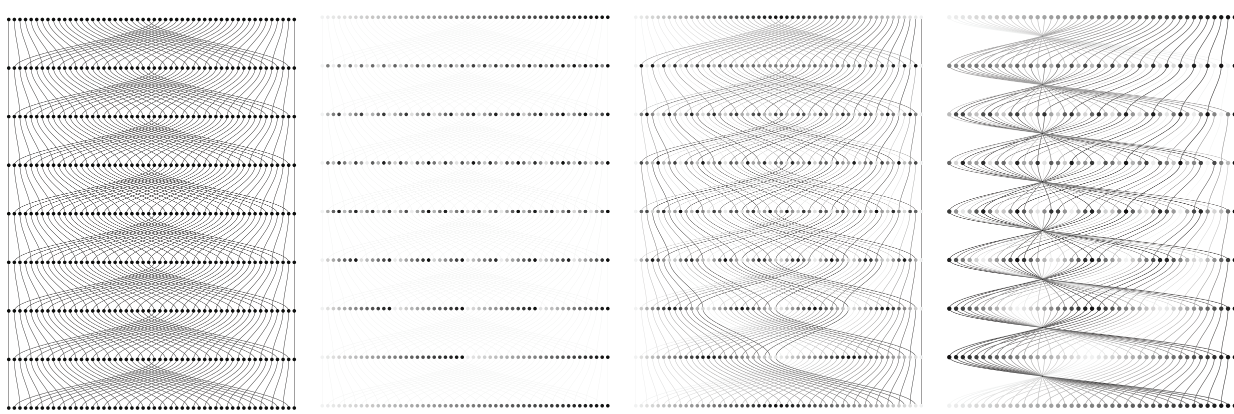

6 *in-shuffles with* 8 *cards*


4 *two-card cuts with* 8 *cards*


4 *milk shuffles with* 8 *cards*

Choices for Visualizations

While pleasing to look at in their own right, we can do much more in terms of visualizing the underlying structure of the shuffles. Two obvious things we can do is to color the dots and the curves in various ways. There are numerous possibilities for the size, shape, and color of the dots and curves, and these are just a few:



(a)
8 *out-shuffles with* 52 *cards*
dots: black
curves: black

(b)
8 *out-shuffles with* 52 *cards*
dots: light-to-dark
curves: faded

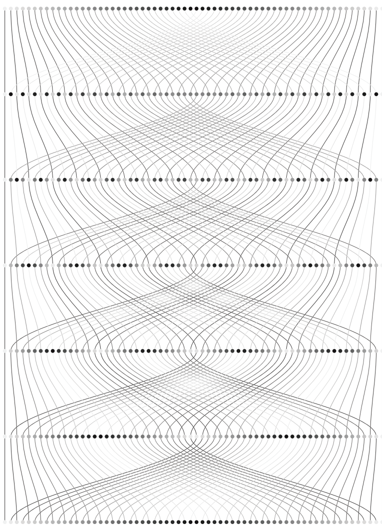
(c)
8 *out-shuffles with* 52 *cards*
dots: light-dark-light
curves: light-dark

(d)
8 *milk shuffles with* 43 *cards*
dots: light-dark
curves: light-dark

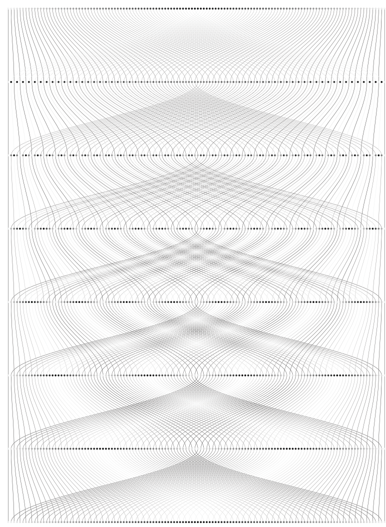
In all of these examples, the original order of the cards is restored after eight shuffles. In (a), (b) and (c), the shuffles are all perfect out-shuffles with 52 cards. For comparison, the shuffle in (d) is a milk shuffle with 43 cards. In (a), there is no particular coloring; consequently, all we see is an identical permutation being repeated eight times. In (b), the dots are colored from light to dark, while the lines are faded to give emphasis to the dots. The coloring of the dots makes it easier to see what the perfect out-shuffle actually does to the cards. For example, we can easily see the intermediate orderings and that the original ordering has been restored after eight times. In (c) the dots are colored from light to dark and *back again*, which is motivated in the next section, while the curves are colored from light to dark. In this way, it becomes apparent that the cards are being *weaved* together. In (d), both the dots and the curves are colored from light to dark.

Visualizing the Stay-Stack Principle

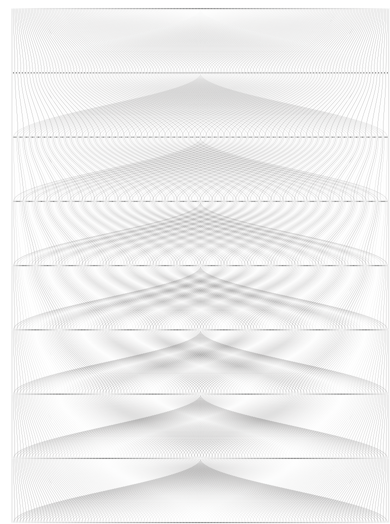
There are many interesting mathematical facts about card shuffling. One is the so-called *stay-stack* principle: If a deck of cards is arranged such that each card n from the top is *matching* the corresponding card n from the bottom, then the deck is said to have *central symmetry*. The stay-stack principle is that any number of perfect in- or out-shuffles preserves this central symmetry. The following three diagrams illustrate repeated perfect out-shuffles until the original order of the deck is restored. The dots are colored from light to dark and back again. Each curve is colored as the opposite of its starting point. Observe that all the diagrams have a left-right symmetry due to the initial coloring and the stay-stack principle.



6 *out-shuffles with* 64 *cards*



7 *out-shuffles with* 128 *cards*



8 *out-shuffles with* 256 *cards*

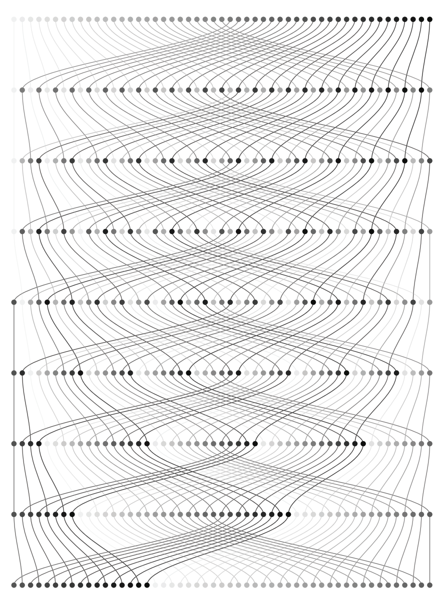
Forcing a Card to the Top

It is now a well-established fact (see [1] for details) that any card in any deck can be forced to the top by an appropriate combination of perfect in- and out-shuffles. The history of this problem goes back to the magician and computer programmer, Alex Elmsley, who in 1957 published a method for forcing the *top card* to any position in a deck. See [2] for a fun typographical application of this principle.

Let us here illustrate another property, which has a similar feel to it, of perfect shuffles: In an deck of N cards, where N is odd, if $2^k \equiv 1 \pmod{N}$, then *any* sequence of perfect in- and out-shuffles corresponds to *cutting the deck* somewhere. (For details, see [3].) For example, in a deck of 51 cards, any combination of eight perfect in- and out-shuffles corresponds to a cut. Look at the illustration to the right. Here, the sequence

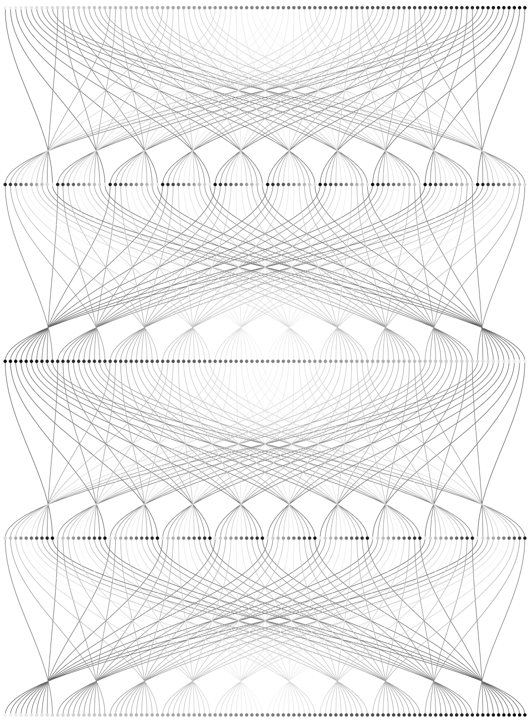
out-out-out-in-out-out-out-in

corresponds to a cut of 17 cards from the bottom of the deck. It is no coincidence that $17_{10} = 00010001_2$.

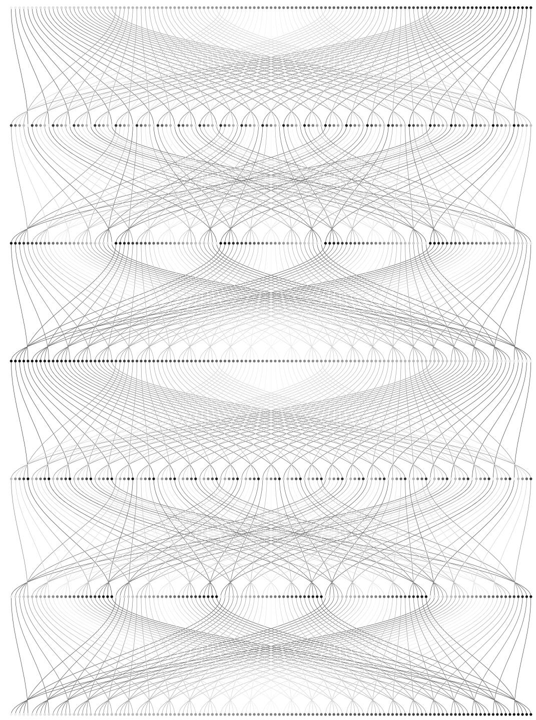


Generalizations

We have only looked at *even* decks of cards and *two* piles in this short paper, but many card shuffles generalize to any number of cards and any number of piles. Here are two such examples that I like:



4 *in-shuffles with* 100 *cards and* 10 *piles*



6 *in-shuffles with* 125 *cards and* 5 *piles*

Conclusions and Acknowledgments

The inspiration for this work came from attending a talk by Perci Diaconis last year. I wanted to know if I could *understand* the mathematics of card shuffling better through appropriate visualizations, so I programmed a first sketch in Processing[1]. My motivation was to be able to *see* some of the underlying structure that would be hard to see otherwise, much like with fractals and other computer-generated art. I find the process of visualizing and experimenting with mathematical structures through computer code, rewarding and exciting.

I view these permutation diagrams as examples of how code can be utilized in order to visualize, and make tangible, mathematical structures, and also uncover some of that underlying beauty and complexity that often follow from very simple assumptions, in this case specific permutations corresponding to card shuffles.

References

[1] Persi Diaconis and Ron Graham. The Solutions to Elmsley's Problem. *Math Horizons*, 14(3):22–27, 2007.

[2] Erik D. Demaine and Martin L. Demaine. Juggling and Card Shuffling Meet Mathematical Fonts. In *Connections in Discrete Mathematics: In Honor of Ron Graham's 80th Birthday*. Cambridge University Press, to appear.

[3] S.B. Morris. *Magic Tricks, Card Shuffling and Dynamic Computer Memories*. MAA spectrum. Mathematical Association of America, 1998.

[1]https://processing.org/

Mathematics and Art: The ELHP

Adam Atkinson (ghira@mistral.co.uk)
Presented at G4G13 April 2018

Some people like to hear about mathematics being used to address real-life problems. I am going to claim that the problem I describe in this article is real-life because it arises from a conversation between two non-mathematicians.

Specifically, one of my sisters-in-law did an art degree, and as part of a project she did for this she visited Sardinia to interview the sculptor Pinuccio Sciola. At least some of his works are quite large, by which I mean maybe 3m high or more, based on things I see on the web. During their conversation, he said something about wanting to install one of his sculptures on a named mountain somewhere in the Catania area, for the benefit of the residents. I don't know his exact words, but my sister-in-law found this remarkable enough that she reported it to me and other members of the family.

It may be important to note that my sister-in-law lives in a small town near Catania, and this may be what prompted him to say this. It seems entirely likely that he had not spent any real time looking into this idea. Indeed, I am told that when he later visited Catania he immediately realised that his idea was probably unrealistic.

Let's use maths, and some other disciplines, to consider his idea, pretending for the sake of discussion that he or someone else really does want to go ahead with it. You may notice that even though Sciola named the mountain, I have chosen not to do so. I shall do this later.

For starters, let's have a look at the mountain. I took this photograph from just outside Catania airport.

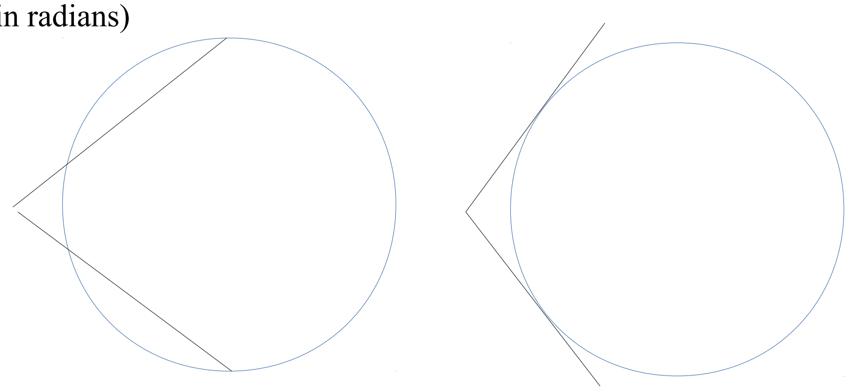
Catania itself is mostly nearer to the mountain than the airport is, but if we're going to build a sculpture or statue on a mountain to be seen from a city with an international airport, we might as well make it so that people arriving from overseas can appreciate the statue as soon as they arrive. Apart from anything else, when you're in town, buildings are often between you and the mountain.

I don't know that Sciola said where on the mountain he wanted to place a sculpture, but it seems as though if you're going to do this at all you might as well put the sculpture right on top of the mountain.

Our first discipline might be geography: how far is the peak of this mountain from Catania airport? There is a "measure distance" feature on Google Maps, and drawing a line from where I took the photo to a reasonable guess at where the peak is, I get 32km. There's also an issue of vertical separation, since the mountain is 3329m tall. There are plenty of higher mountains in the world, but that's not bad. (From my father-in-law's house to the peak is about 26km horizontally.) The airport is very near the coast, and my eyes aren't that far above the ground. Let's pretend the photograph was taken at exactly sea level. We're not going to be pretending anything we do has more than about 1 significant figure anyway. Using Pythagoras we get straight line distance as $\mathrm{sqrt}(3329^2+32000^2)=32173$m. Let's just say 32km.

Since we can see the mountain, we could at least in principle see things on it. But how big would they need to be? We're probably asking how large a solid angle in steradians the sculpture needs to subtend, or what its angular diameter needs to be in degrees or radians. If we assume the sculpture to be approximately spherical, we don't need to worry about foreshortening caused by looking up at it either. Since we know we can see very distant stars in the night sky, and they have extremely small angular diameters, it's clear that there isn't really a minimum size our sculpture could be to be visible at all: if it gives off enough light we'll be able to see it no matter how small it is.

Here we see that the angle subtended at the eye by the diameter of a sphere (perpendicular to line of sight) is smaller than the angle subtended at the eye by the sphere as a whole. Clearly in this example both angles are quite large and the difference between them is substantial. With much smaller angles, as we will be discussing, the difference is much smaller since x, $\sin x$ and $\tan x$ are all very similar for small x. (With x in radians)



When I calculate the sizes of statues at given distances I will just use $2\pi r(\text{angle}/360)$ (with the angle in degrees). For the very small angles we will be talking about, this is good enough. I am ignoring the fact that the distance to the centre of the statue increases as the statue gets larger. Taking the height of the mountain into account did not make much difference and for our purposes we can for now ignore the extra distance induced by changes in statue size. And the fact that the Earth is not a perfect sphere.

But would a sculptor be satisfied with a sculpture that is for all practical purposes a single pixel? I'm not an artist, so don't really know, but it seems unsatisfactory. I can see that an artist might choose the colour of a single pixel with great care. The intensity and colour could even vary with time if we wanted something more than a static dot. However, let us suppose that this is not what Sciola had in mind.

At this point we need to know something about human physiology. About how big should the smallest details of the sculpture be for people to be able to see those? It appears (see e.g. http://www.bbc.com/future/story/20150727-what-are-the-limits-of-human-vision) that the finest pattern the human eye can distinguish is about 120 lines per degree. So single dots are half a minute of arc. At 32km, half a minute of arc works out to be about 4.7m. Let's say 5m. We could start to think in terms of pixel art with 5m pixels but perhaps this approach is wrong.

We could instead consider existing statues or other items which are viewed from some distance away, and where more than their mere existence can be perceived, and ask ourselves what their angular diameters are when viewed from those distances. If these statues or other objects are approximately spherical, so much the better. We can of course consider things such as the Angel of the North, Christ the Redeemer and so on but the Moon seems to be a very good candidate. It's approximately spherical, its angular diameter is about half a degree, and "about the same size as the Moon" would not be an outrageous starting point for what we might want a statue to be. If we want some kind of "moon unit" like this to use in comparisons, we could call it the Zappa. Half a degree, at 32173m, means a diameter of 281m.

What should our statue be of? One might think that the patron saint of Catania, Saint Agatha, would be a candidate, or the elephant, the symbol of Catania. The problem with these is that the top of the mountain can be seen from other, closer, communities who might feel that this was unreasonable. And Saint Agatha, as usually depicted, is not approximately spherical.

I submit, then, that a solution which should be acceptable to all is to build a statue of my stuffed hedgehog, Herisson. (Photo by Adam Atkinson)

Herisson is round enough that I think we can treat him or her as approximately spherical.

And so we have the Extremely Large Herisson Project: we wish to erect a statue of Herisson on top of a mountain 3.3km tall, 32km away, so that it appears to be about the same size as the moon. Let's say that we want Herisson's largest dimension to subtend the same angle that the Moon does.

As seen above, this works out to be about 281m. This is of course a lot. However, looking for large statues we find that the Spring Temple Buddha in China is 128m tall, on a 25m base. Our proposal is not fantastically larger than the largest statue in the world. Alternatively, we could settle for the Extremely Large Herisson being a little less than one Zappa in size, perhaps.

There are non-mathematical considerations, however. Merely doing calculations with similar triangles ignores the problem of haze. When in Catania recently trying to take a photograph of the mountain, 5 days out of 7 it was behind haze, fog, cloud or similar. Being able to see anything at all at 32km, especially when it is above common altitudes at which clouds are found, is problematic.

One is also driven to wonder how planning permission for this project would work. On Wikipedia we find this image, (By Skyluke - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=1961715) which purports to show the town boundaries in the Catania area.

The mountain peak is the point in the top right where multiple sectors meet. It might be necessary to obtain planning permission from all the towns whose areas meet at the summit. Might it be necessary to get two lots of planning permission from the town whose area meets the summit twice? Of course, since much of the mountain is a national park, there could be further complications. I may ask at least one of the town councils about this project at some point. I have in fact written to one town council, that of Nicolosi, asking about this. At the time of writing, in late June 2018, I have had no reply.

Building a 281m statue on top of any 3329m mountain would be a challenge, but in this case there are extra difficulties: The mountain under discussion is in fact Mount Etna, one of the most active volcanoes in the world. Google can find plenty of images of spectacular events on and around Etna.

One experiment we can do is to view normal-sized Herisson at suitable distances to see if one Zappa does indeed seem about right, and also try to see if when his or her angular diameter is half a minute he or she is indeed visible at all. (He or she is white, so against many backgrounds might stand out enough to be seen.) A very long tape measure would of course be needed.

Here I am on a field trip to Catania (photo: Annamaria Cucinotta), holding up a copy of Chalkdust magazine with Etna in the background. Ideally we should have made sure the distance to the camera and the size of the magazine were such that it looks the same size as the Moon would on the mountain behind me. This photo serves to illustrate a way we could decide what a decent size of statue would be. A stepladder would clearly have made it possible to hold the magazine so that it seemed to be on top of Etna, but one was not immediately to hand.

On a later field trip it seemed useful to examine the proposed construction site itself. On the left I am in front of the main crater with a normal-sized copy of Chalkdust. The photo on the right has some people in it to give a sense of the scale of the place. (Both photos by Gunther Schmidl) The main thing I learned on this trip was that I never want to do it again. Making it all the way up to the crater is quite hard. Transporting construction materials for the sculpture would be a major undertaking.

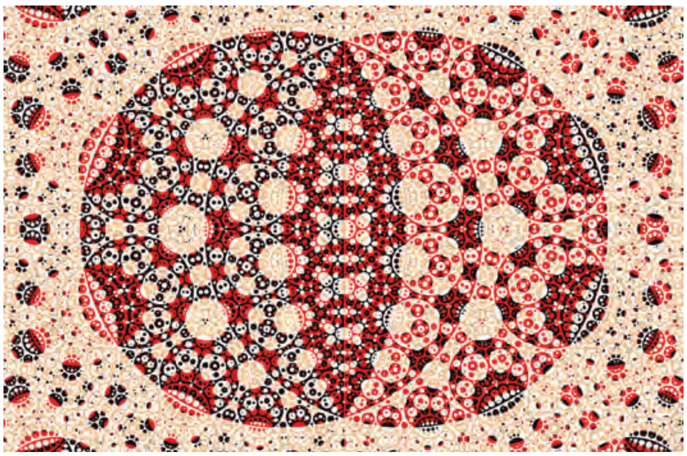
But why stop at Catania? The city of Messina is about 70km away. And if we calculate the distance to the horizon from the top of Etna we discover that it's about 206km. 160km gets us to Palermo. 203km gets us to Catanzaro in southern Italy. Most of Sicily is within 206km of the peak, but we can't quite get to Trapani or Marsala on the west coast of Sicily. However, the sculpture would be so big by this point that its feet on the mountain would be invisible but some of the sculpture could still be seen. Still, I think we have to insist that the feet be visible: if we make a tall thin statue big enough of course the top can be seen from almost half the planet, and some part of a large enough spherical statue can be seen from almost anywhere on the planet (though the statue may need to be larger than the planet itself). Also, if your eyes are higher than sea level you can some extra range from that. One photograph I have seen on-line of Etna from Catanzaro was taken from a hotel balcony. It seems possible that from a tall enough building in Trapani or Marsala, the top of Mount Etna might just be visible if there's nothing in the way.

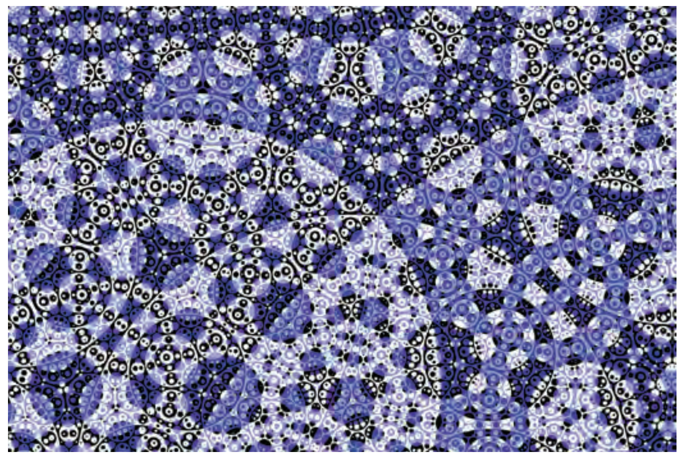
Note that I am completely ignoring refraction here. In real life, refraction could make a difference, and as the wonderful article by David E.H. Jones (a.k.a. Daedalus) in "New Scientist" https://bit.ly/2lzswd8 observes, if we replaced the Earth's atmosphere with sulphur dioxide or reduced the Earth's radius we could in principle see as far as we wanted. But both of these options are well outside the scope of the ELHP.

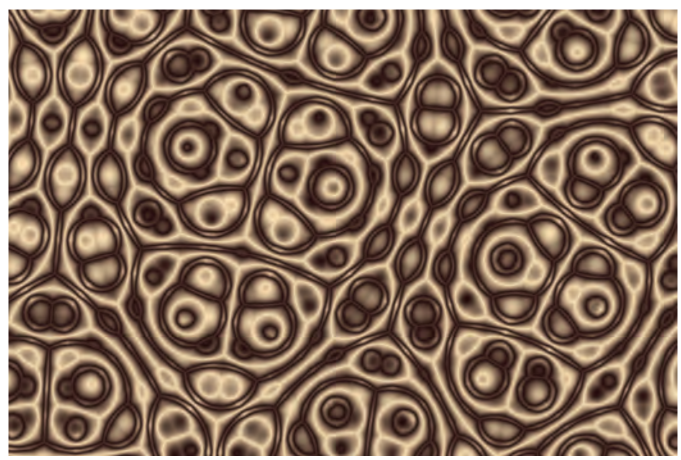
Putting 206km instead of 32km into our formulae, we get "pixels" of about 30m and statue size 1798m. At this point we have a Prodigiously Large Herisson Project, and it seems quite impractical. And there's the question of which way he or she should be facing. Which cities get to gaze upon Herisson's gigantic backside? Gunther Schmidl (personal communication) suggests making the statue rotate about a vertical axis. This is clearly the way to go.

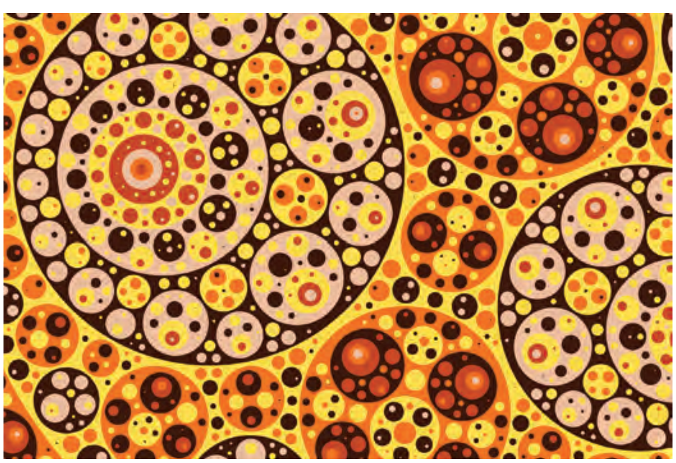
Cross sections of three dimensional hyperbolic tilings *by Vladimir Bulatov (bulatov.org)*

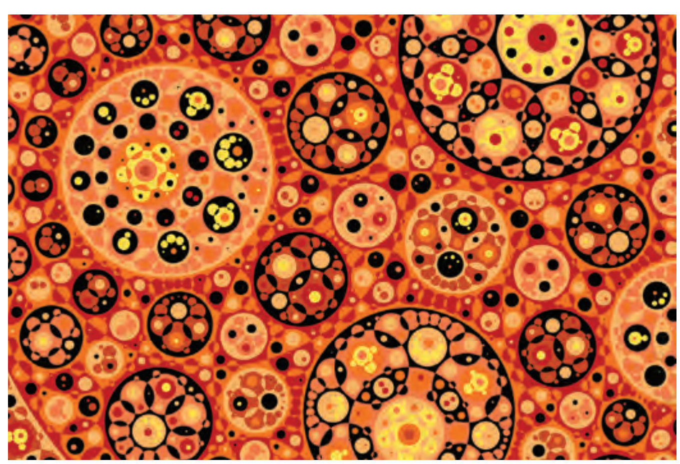
Three dimensional tiling of hyperbolic 3-space is difficult to visualize. In the standard models of hyperbolic space like Poincaré ball or upper half space the tiles become very small very fast near the boundary of the hyperbolic space. The complete tiling is actually quite boring – just a round ball or half space. Taking cross section of the tiling with hyperbolic planes gives picture similar to images of two dimensional hyperbolic tiling. However there is very special kind of surface in the hyperbolic space - horosphere. It is surface of zero curvature and it's intrinsic geometry is geometry of euclidean plane. All horospheres are equivalent to each other. The horospheres in the ball model are spheres tangent to the ball boundary. Horospheres in the upper half space model are spheres tangent to the boundary plane. Especially convenient are horospheres, which have tangent point at infinity. These are just planes parallel to the boundary plane. The cross section of some three dimensional hyperbolic tiling by such horosphere are illustrated here. Nice property of such cross section is the constant scale of the pattern. It is consistent with the euclidean nature of the horosphere geometry. Another nice property is that the pattern is not periodical and does not repeat itself exactly and also keeps the similar uniform appearance everywhere.









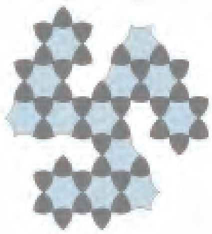
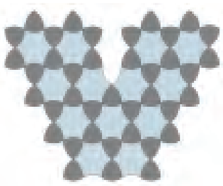
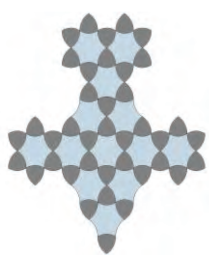


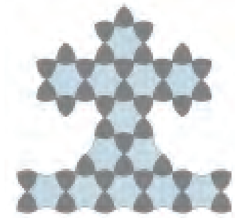
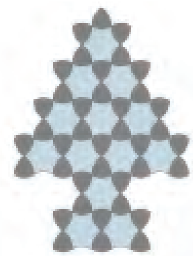
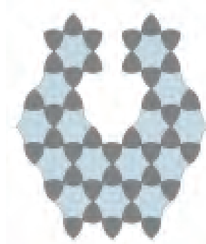
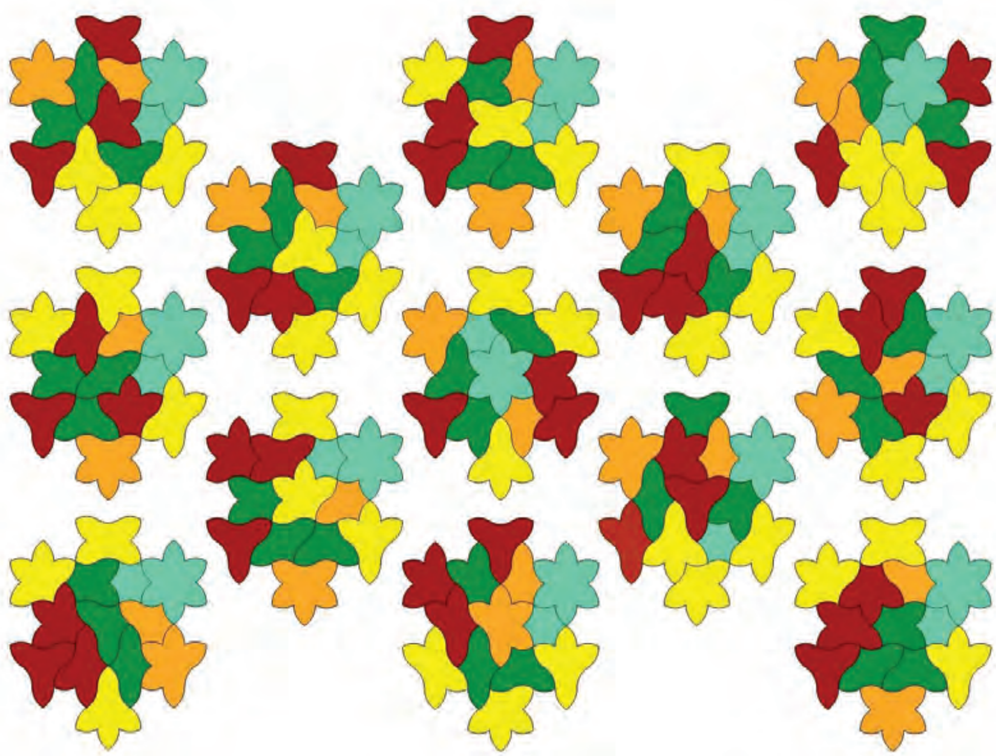
Geometry in Five-Dimensions: Building Quasicrystals from Penrose Tiling

Debora Coombs

Cookie Jar™

Designed by Michael Dowle, expanded by Kate Jones

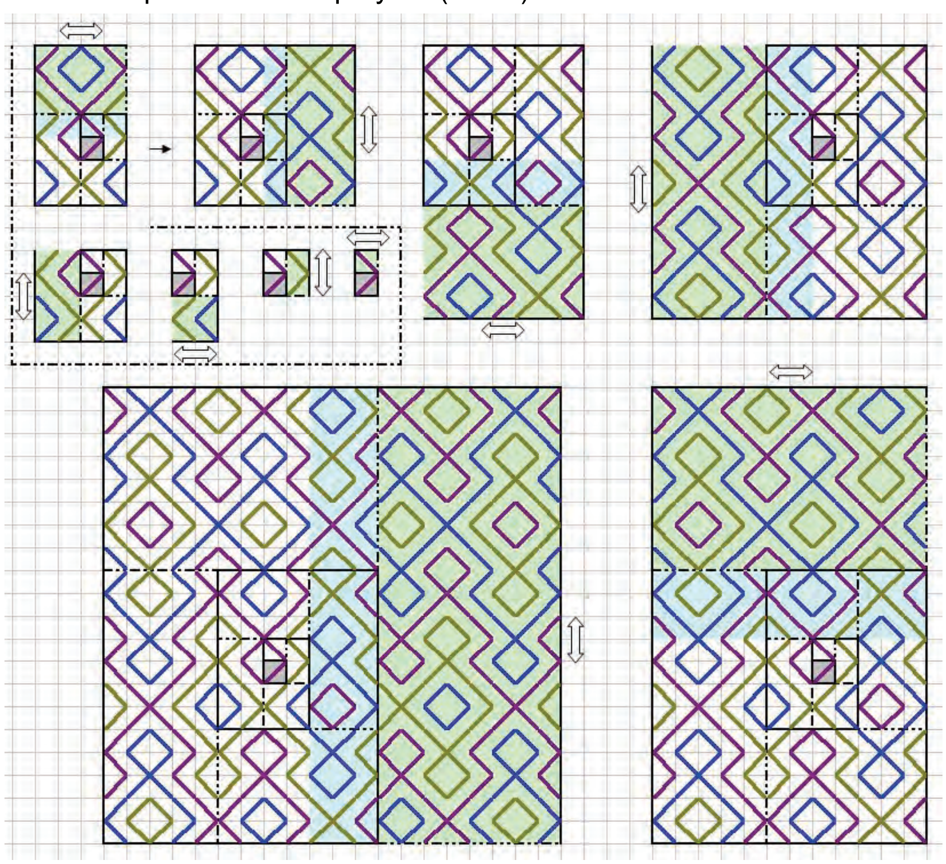
Unique Three-Color Colorings of Aperiodic Tilings

David Michael Greene

For many years I've been fascinated by a unique three-coloring of an aperiodic tiling discovered by Robert Ammann -- the golden-bee tiling, or "Ammann's A2" in Grünbaum and Shephard's <u>Tilings and Patterns</u>. The G4G Exchange book seems like a good place to share a few pictures. Please feel free to continue the investigation from here!

Start with a graph consisting of a unit square with two of the opposing corners connected. This will be called "stage 2" of the graph. Stage 0 is a point, stage 1 is a unit horizontal line -- it's simplest to start with stage 2. It doesn't matter at this point which initial diagonal is chosen.

Each stage N+1 of the graph is composed from two copies of stage N. The second copy should be reflected in the Y axis and translated Fib(N\2) units downward in relation to the first copy. The two copies will overlap by Fib(N\2-1)-1.



N\2 denotes integer division by 2. Fib(0)=1, Fib(1)=1, Fib(N)=Fib(N-1)+Fib(N-2).

After the copy, reflection, and translation operations, rotate the resulting larger graph 90 degrees clockwise to obtain stage N+1. For even N, the stage-N graph will be a square of side Fib(N/2+1)-1. For odd N, the graph will be a rectangle Fib((N+1)/2)-1 units wide, by Fib((N-1)/2)-1 units tall.

The first several stages of composition of the resulting "Golden Graph" are shown above, in a spiral starting from the upper left quadrant of the figure.

In 2002 I found a way to generate the above graph, and thus the associated three-color pattern, fairly directly from two self-generating sequences related to the "Golden String" sequence:

S(0) = "0", S(1) = "1", S(N) = S(N-1) & S(N-2).

Here "0" corresponds to a square with (say) NW and SE corners connected, and "1" is a square with NE and SW corners connected. Determining the exact recursion -- a different starting string for each of the two dimensions -- and applying three colors to the resulting graph is left as an exercise for the reader.

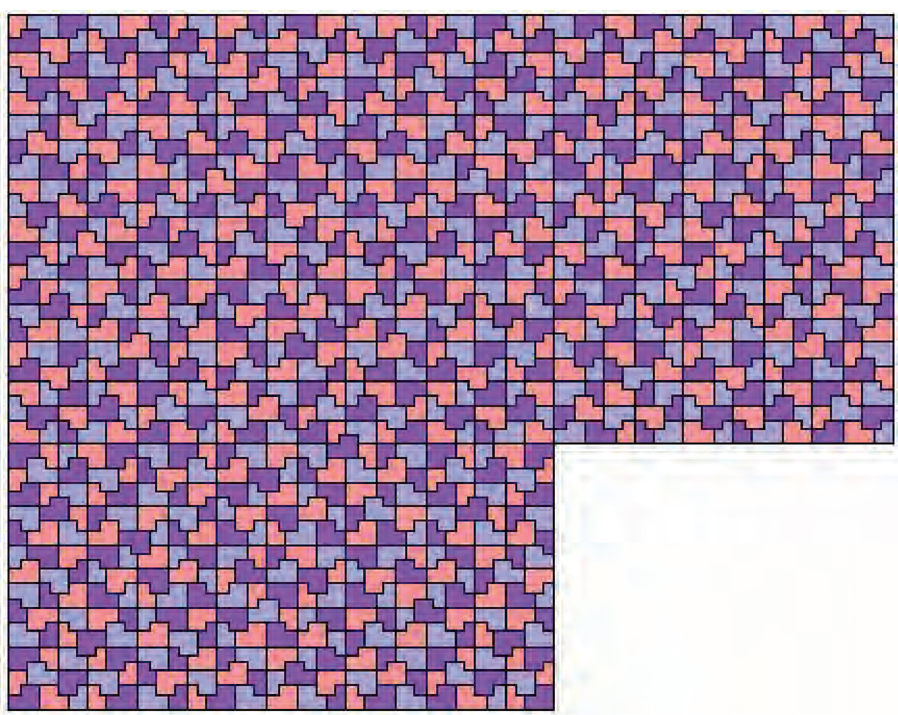
There is really only one color pattern in each direction. If the top of a column or the beginning of a row is a given color, any column or row starting with that color will match the colors of the original column or row, all the way to the end. Columns and rows starting with other colors can be duplicated by applying the same color shift (modulo 3) to the original column or row.

An alternate way of obtaining this same colored graph is to start with a single Golden-B (or "Golden Bee") tile, corresponding to stage 0 of the graph. See

http://www.meden.demon.co.uk/Fractals/golden.html

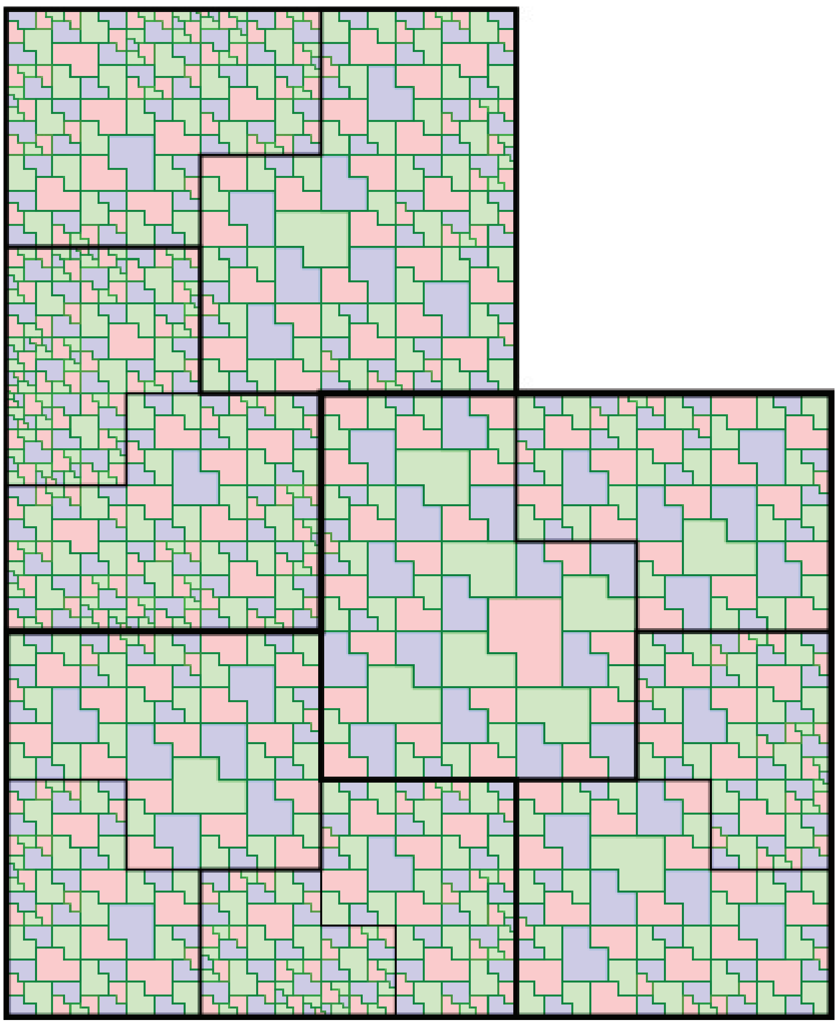
or Grünbaum and Shepard's book Tilings and Patterns for more information about this tile.

To produce each stage N+1 from stage N, cut each of the Fib(N-1) _larger_ copies of the tile into two similar small copies. At any stage, the resulting tiling seems to be uniquely three-colorable, as long as you include the additional requirement that all three colors must be used wherever four tiles meet at a corner... and that three-coloring matches the unique three-coloring of the Golden Graph described above.



The "Golden Graph" can be produced by replacing all tiles with nodes, and drawing lines between each pair of neighboring tiles that have the same color. (Each interior tile has eight "nearest neighbors"; note that they may not be actually adjacent). Alternatively, drawing lines between pairs of neighboring tiles whose colors do <u>not</u> match is equivalent to rotating all the unit-cell diagonals in the graph by 90 degrees.

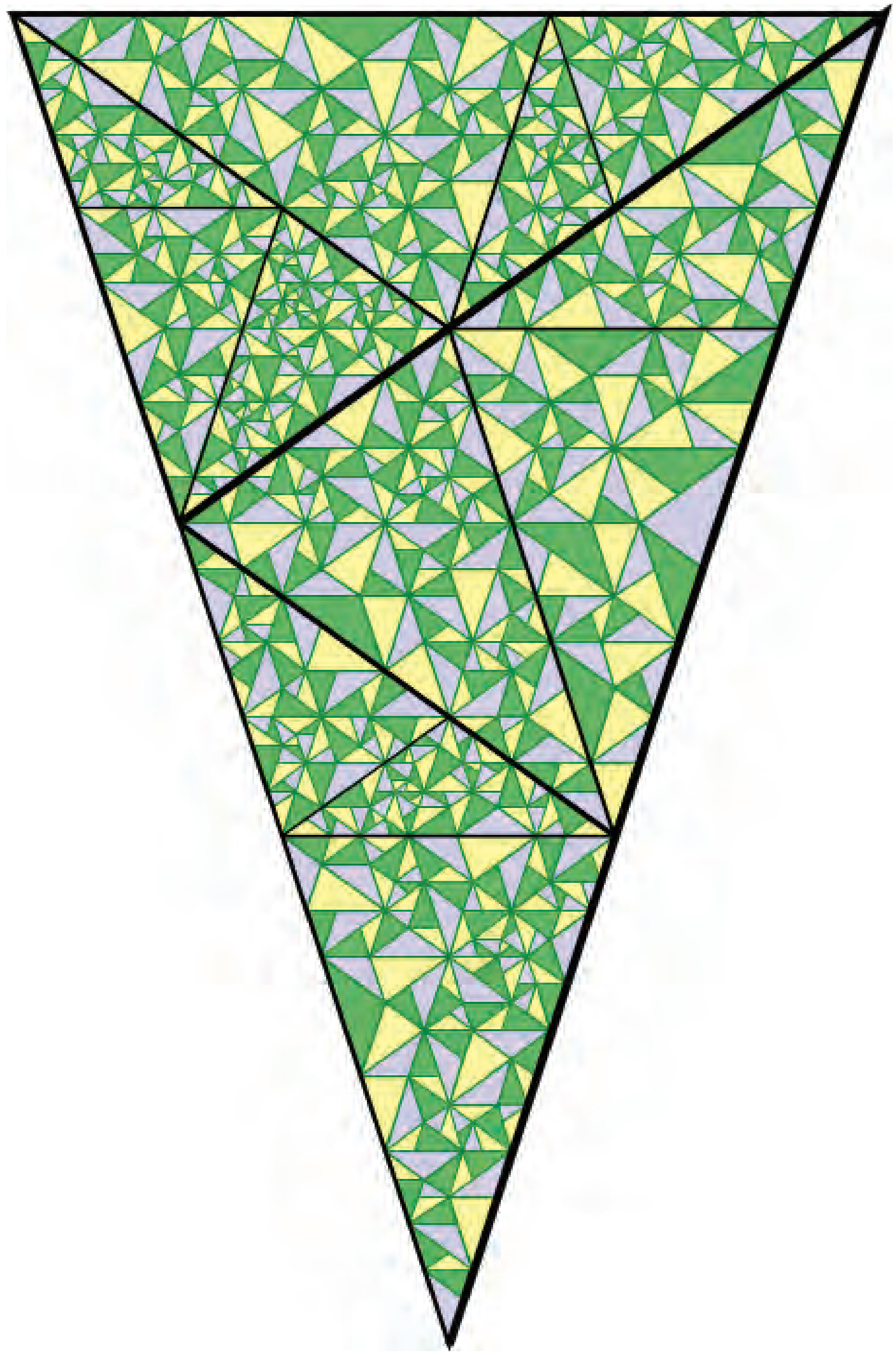
Another three-colorable tiling using the Golden B tile can be obtained by cutting *every* tile in each stage into its two similar sub-tiles, instead of just the larger tiles in each stage. Stage N of this tiling has 2^N tiles instead of Fib(N), with tiles of N different sizes.



With 0, 1, and 2 as possible colors, if the color of the *larger* sub-tile is incremented (mod 3) after every decomposition, a unique three-coloring is the result. The smaller subtile's color remains the same after each cut.

There is no need in this case for the additional restriction that all three colors must be represented at each corner. Only two colors will appear at some corners, but no two tiles of the same color will ever meet at an edge.

This kind of recursive coloring rule also works on another aperiodic tilings, apparently unrelated except for another appearance of the Golden Ratio.



Iterated diissections of Robinson triangles (half Penrose rhombs) can be three-colored in this way.

Here again, stage N of this tiling has 2^N tiles instead of Fib(N), with tiles of N different sizes.

With 0, 1, and 2 as possible colors, if the color of the *smaller* sub-tile is incremented (mod 3) after every decomposition, a unique three-coloring is the result. At each stage, the larger sub-tile's color remains the same.

A two-size Robinson triangle dissection can be created, analogous to the two-size Amman A2 tiling in that only the larger of the two Robinson tiles is dissected at each stage of the inflation process.

However, there appears to be no unique three-coloring equivalent to the two-size Ammann A2 tiling case.   A trivial two-coloring is possible, simply because such dissections can only have an even number of triangles meeting at any vertex.

Are there other similar examples of unique recursive three-colorings of aperiodic tilings?

A Recipe for a ‘bola Honeycombs

David Hall

David.42.Hall@gmail.com

Abstract:

A hexagonal grid and simple integer addition can be used to generate a set of coordinates which all fall on a three-dimensional surface called a paraboloid. The grid generates the vertices of affine hexagonal facets which bound an infinite polyhedron that I have dubbed the "Parabolahedron" (hence 'bola in the title). The entire parabolahedron is completely determined by the choice of four "seed numbers" from which the entire polyhedron is derived. An infinite variety of these parabolahedra may be generated by choosing different "seeds".

Once completed, a grid can be used directly for modeling a paraboloid by stacking anything from coins to Honeycomb cereal in stacks corresponding to the values in the grid. The result is a paraboloid shaped "bowl".

Since the process of generating the honeycomb grid requires only integer addition, it is a suitable puzzle for students in early elementary grades. Any of several pre-calculated grids could be provided to allow even younger students who are just learning to count to generate variants of the "bowl" shape by stacking simple objects such as coins. At the same time, the geometric progression of the grid can be used in more advanced studies of topics including averages, slope, volume, exponential growth, symmetries, conic sections, quadric surfaces, vector addition and affine transformations.

Introduction:

A hexagonal grid is the basis for a simple, yet intriguing puzzle presented here in two parts. The first part is the 2D puzzle itself which is reminiscent of Pascal's Triangle both in its simplicity as well as the hidden gems it reveals. The second part takes advantage of the printed puzzle to make some actual 3D models of the shape. In closing, I’ll offer some ideas for use of this puzzle in various classroom environments.

The Honeycomb Grid:

The grid consists of any number of identical hexagonal cells laid out like a honeycomb as shown in Figure 1. The hexagons need not be regular. That is, it is not necessary for them to all have all edges the same length or all angles identical, but it is necessary for them to have all three pairs of opposite sides parallel to the diagonal between each of them. The following examples uses regular hexagons to introduce the concepts. The vertices of the grid where three hexagons join are indicated in red. The centers of each cell (called faces) are indicated in green and the edges of each cell are bounded by a pair of lines which will occasionally be designated “slope”.

An arbitrary vertex is chosen. It’s shown in bright red in Figure 1 surrounded by three edges highlighted in yellow. The grid is populated by selecting any numbers for the chosen vertex and the three adjoining

yellow edges, then calculating values for the remaining vertices, faces and edges according to the following rules.

1) The value on any edge is copied to the opposite parallel edge as well as each side of the center between the green and red points as shown in Figures 2a, 2b and 3.
2) The values of two edges with a common vertex are added to produce the remaining edges, which is then copied to the two remaining internal positions.
3) The initial red cell is added to the adjoining edge values to produce the values of the adjacent vertices which are in turn added to their adjacent edge to produce the 4th and 5th vertex.
4) Either the 4th or 5th vertex can be added to its adjacent edge to derive the final vertex, which is the one opposite the first one.
5) The value of the face, indicated in green can be determined in any of several ways, all of which should produce the same value. It is the sum of the vertex closest to the initial red vertex and the "slope" value between that vertex and the middle. It is also the value of vertex farthest from the initial red minus the "slope" between it and the middle. It is also the average of every pair of opposite vertices for that hexagon.
6) If the various methods for calculating the green value in a cell don't all agree, then stop and check your earlier math before continuing to the next cell. Figure 4 shows an example of a completed cell.
7) The new edge value determined in step 2 can now be copied to all applicable positions in the grid as described in step 1 and then repeat the process for any hexagon having a value for its vertex and two adjacent edges.

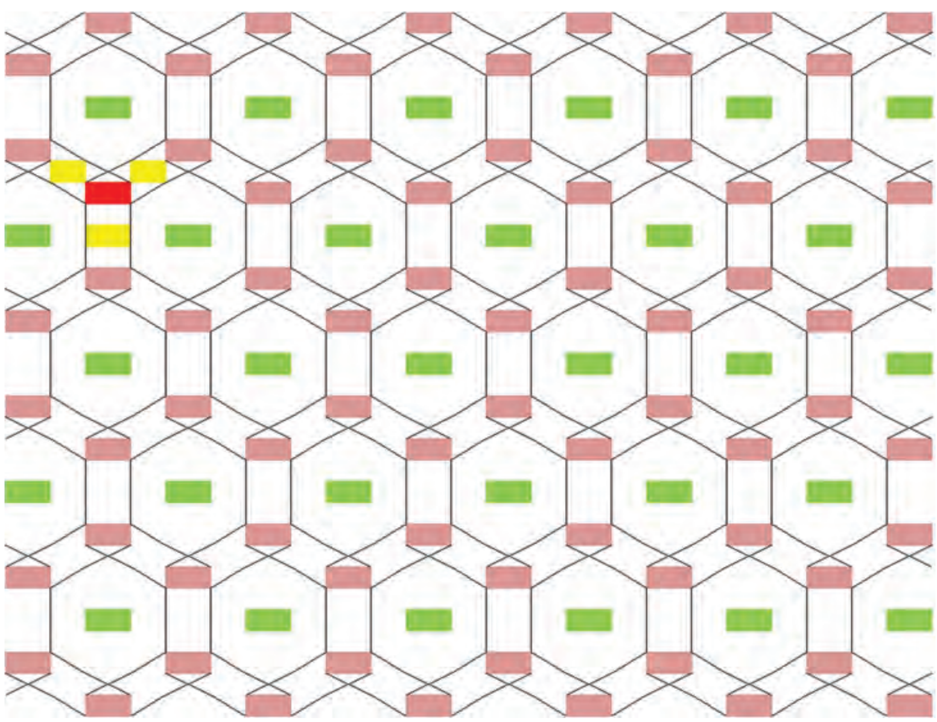
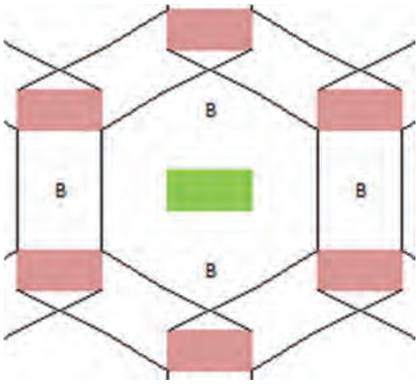


Figure 1: Blank honeycomb grid with an arbitrary starting point highlighted.



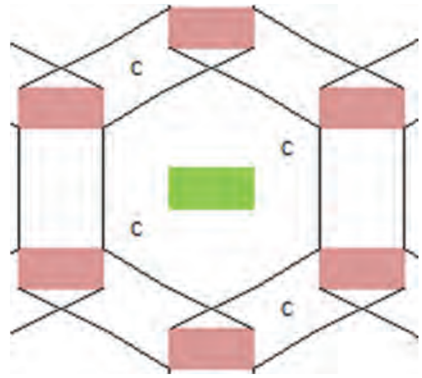


Figure 2a: Slope "B" is copied horizontally. **Figure 2b: Slope "C" is copied diagonally.**

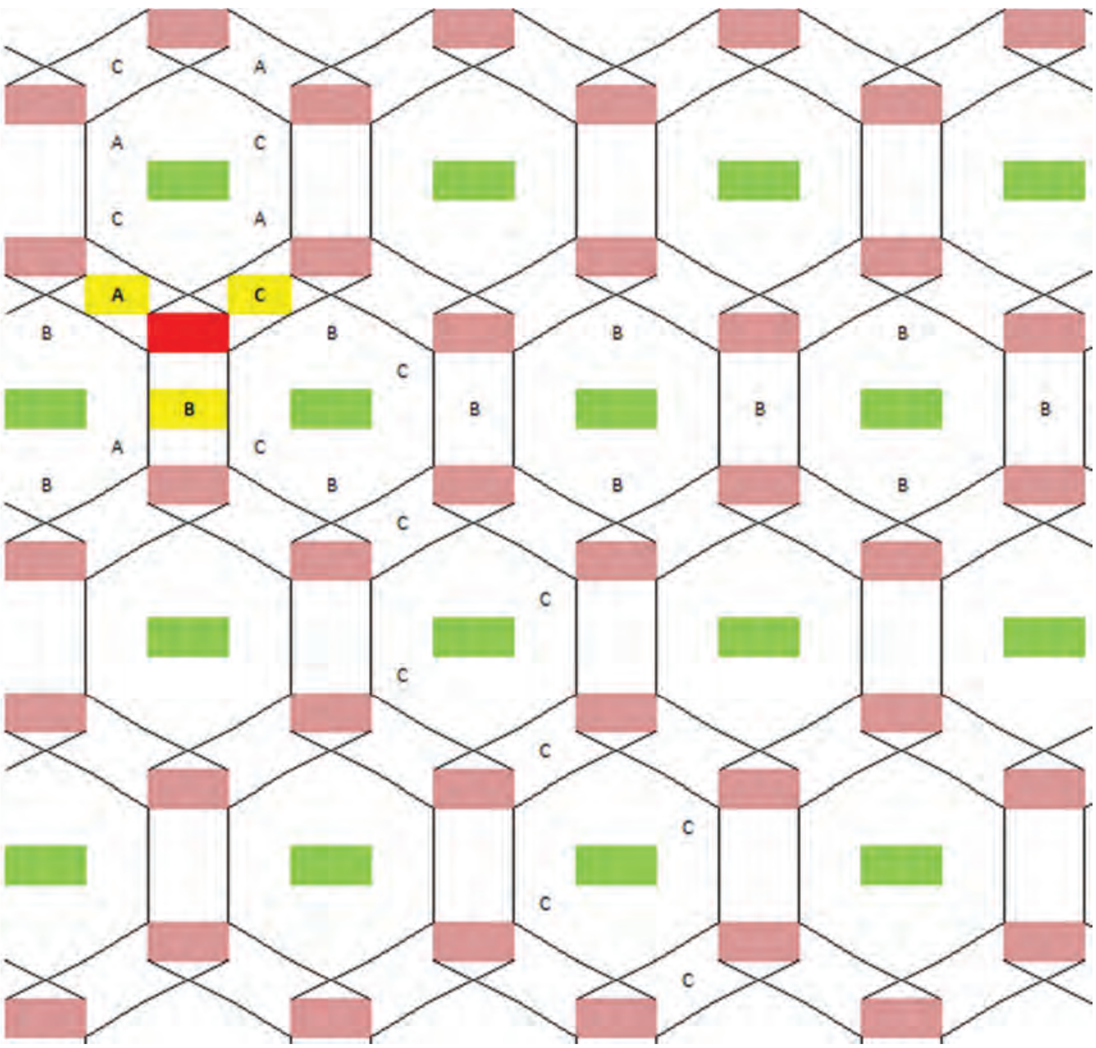


Figure 3: Initial edge values of A, B and C can be copied to the rest of the grid as shown. The actual puzzle would use numbers, not letters in these positions.

Figure 4: Relationship between vertices, edges and middle of each cell.

Here's a step-by step example using the cell on the right side of Figure 4:

1) Values of 2 and 3 are copied from the two yellow edges to opposite edges and internal slopes.
2) 2 + 3 = 5 which completes the edges and internal slope values.
3) 0 + 2 = 2 and 0 + 3 = 3, then 3 + 5 = 8 and 2 + 5 = 7
4) Either 8 + 2 = 10 or 7 + 3 = 10 which is the last vertex of this cell.
5) The green center is 0 + 5 = 5 or 10 – 5 = 5 or average(0,10) or average(3,7) or average(2,8).
6) The cell on the left is solved similarly. The other cells around them are shown partly solved.

<u>What's the point?</u>

Now that we have our grid, we can use the numbers in red to get a list of three dimensional coordinates. The numbers in green and white areas will come in handy later too. Those in white, as you may have already realized, have to do with the slope of a facet. Those in green are important coordinates too, but they don't lie on the surface of the paraboloid. They're useful for calculating volumes too as we'll see.

For now, let's just determine the list of 3D coordinates. It was mentioned earlier that the hexagons don't need to be regular. They just need opposite edges and their diagonals to be parallel. With that in mind, we can use any unit for measuring our grid in the X and Y directions. I prefer to start with 0 in the bright red vertex, then I can simply count hexagons vertically or horizontally. The hexagon in Figure 4 is divided into smaller rectangles. They provide an easy way to count off the X and Y coordinates. Since we won't be concerned with the white slopes as coordinates, we can skip them and just count every other column for the X coordinate and every other row for the Y coordinate, and best of all, we can use the value in the red cells directly as a Z coordinate. Again, the units don't matter. Our X, Y and Z coordinates don't even need the same units as each other.

With this basis for measuring our set of coordinates, the vertices shown in Figure 4, listed from left to right and top to bottom are: (-1, 1, 1), (1, 1, 3), (-2, 0, 4), (0, 0, 0), (2, 0, 8), (-2, -2, 6), (0, -2, 2), (2, -2, 10), (-1, -3, 5), (1, -3, 7). As a side note, it's interesting to point out that, just as 3 points define a circle and 5 points define a conic section (e.g. ellipse or hyperbola), 9 points are needed to define a quadric surface like our paraboloid, and we have calculated 10 points so far. Calculating the formula for our paraboloid is beyond the scope of this article, but it's interesting to note that we already have enough information to do so.

Honeycombs… They're not just for breakfast any more.

But wait, since the dimensions are all independent and can use different units, we can finally start to develop our simple 3D model.

Let's simply use the grid directly without counting off X or Y coordinates, and then we can decide that a unit in the Z direction (perpendicular to the grid) is whatever we have handy. If you have a few dollars in pennies available, then define one unit in the Z direction as the thickness of one penny, then make stacks of pennies right on top of each red vertex. The number in the vertex represents the number of pennies in the stack. If you started with big numbers, your bowl will get expensive. If you started with small numbers, like 1's and 0's, the thickness of a penny won't add up to a very deep bowl, and that's where the Honeycombs come in. Honeycomb cereal is about 4 to 5 times as thick as a penny, so they can be stacked in similar fashion and the bowl will grow more quickly so you'll finally start to see your "bola" Honeycombs. Of course, you're free to experiment with this recipe and substitute just about anything you'd like for the stacks of Honeycombs. The result will still be points on a paraboloid.

Figure 5: Piles of Honeycomb cereal stacked on the hexagonal grid begin to curve upward and form the promised 'bola Honeycombs. Pennies or other items can be substituted to form the stacks. The cereal was laced onto Angel hair pasta to keep them from falling over. Pennies are easier to manage.

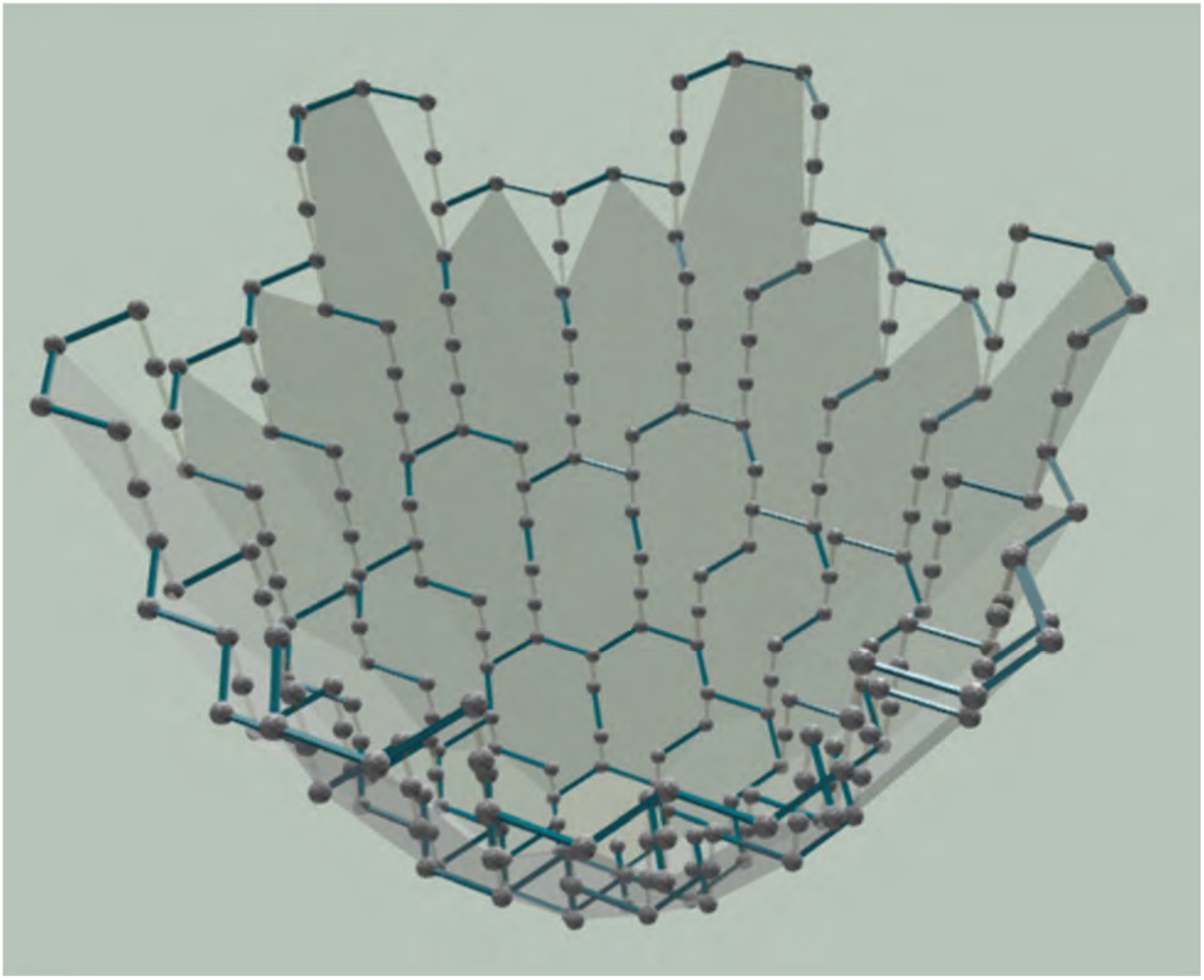


Figure 6: A parabolahedron made by stacking zometool balls and struts. This interactive 3D model is available on Sketchfab at https://skfb.ly/6wXPF.

Consider the alternatives:

At this point, we'll look at the outcomes from choosing various initial values. This will not be an exhaustive study, but it will point out some key characteristics to notice when experimenting with the grid.

Ultimately, the grid is the basis for modeling our paraboloid, or more specifically, a subset of a paraboloid which I've dubbed a Parabolahedron, so let's continue with a few simple definitions.

1) A parabola is a conic section. It's the two-dimensional shape that results from slicing a cone parallel to its surface.
2) If you spin a parabola around its axis of symmetry, you get a paraboloid. It's a three-dimensional shape that you might recognize as a reflector for a lamp. If you slice it perpendicular to its axis, you'll find a circle.

3) Squashing a paraboloid so that the circular cross section becomes an ellipse is still a paraboloid. Unsurprisingly, it's called an elliptical paraboloid.
4) Any version of a paraboloid is a smooth continuous three-dimensional curve called a quadric surface. All the points we've generated with our grid will fall on that surface, but they will not completely fill the surface.

It turns out that each hexagon generates a set of 6 points which, even after being projected into a 3rd dimension, remain coplanar. In other words, each original hexagonal cell in our grid corresponds to an affine hexagonal facet of the parabolahedron.

The initial point has been chosen well off center in these examples to feature the extended growth of the red and green coordinates. Symmetries can be used to enlarge the grid based on calculating and then replicating a subset of the coordinates. Regardless of the first vertex we choose, the initial choice of slopes will determine the orientation of the parabolahedron with respect to the plane of our grid. Let's assume for now that the initial vertex is at 0. Several possibilities are described in the following figures.

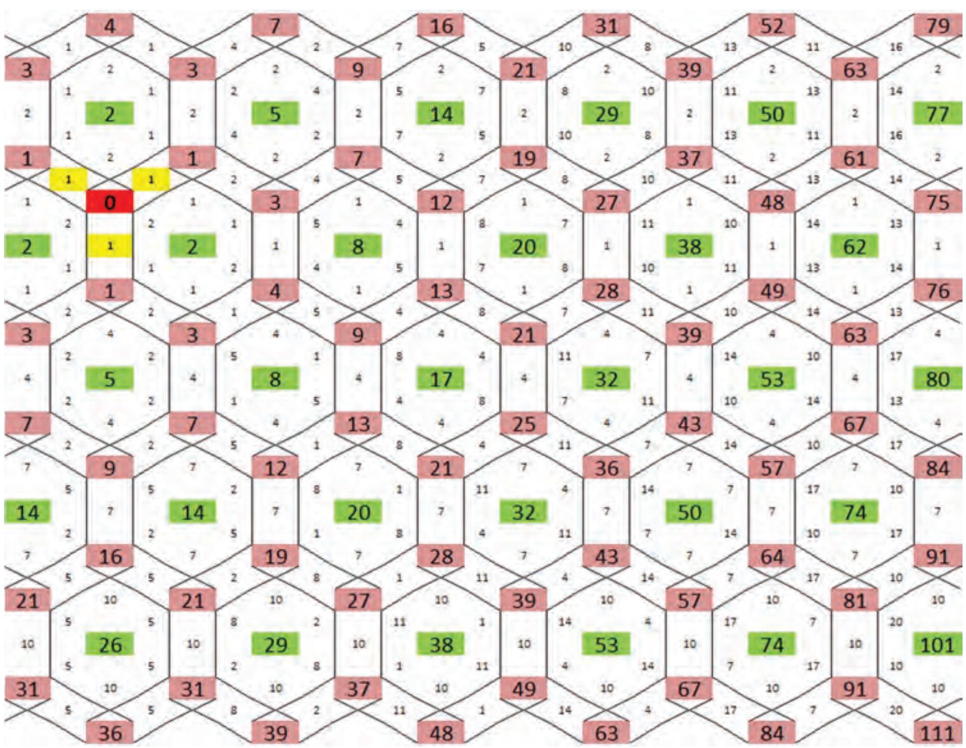


Figure 7: A "vertex first" projection.

1) Three slopes are all the same and not zero. This results in a single vertex touching the plane and the axis of symmetry of the parabolahedron being perpendicular to the plane at that point as in Figure 5.

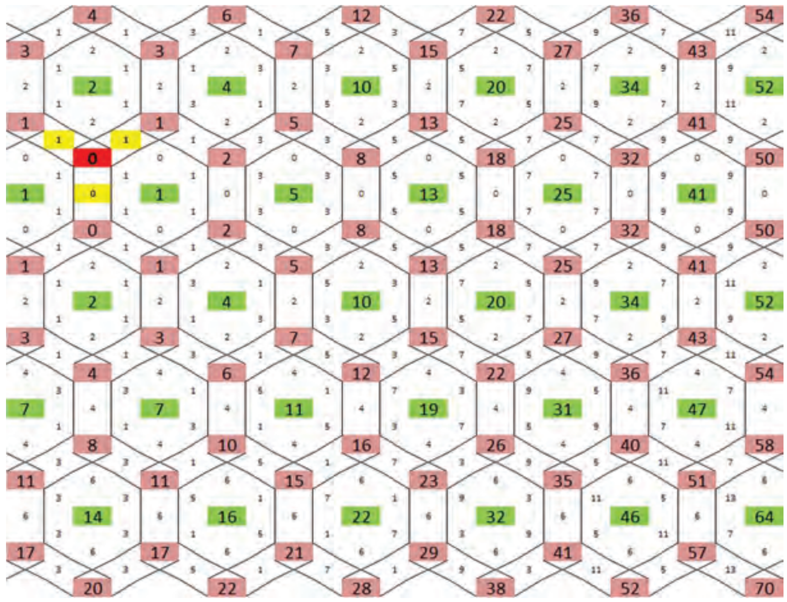


Figure 8: An “edge first” projection.

2) Two slopes are the same and the other is zero. This results in two vertices and their common edge (the one with zero slope) touching the plane and the axis of symmetry of the parabolahedron being perpendicular to the plane through the center of that edge as in Figure 6.

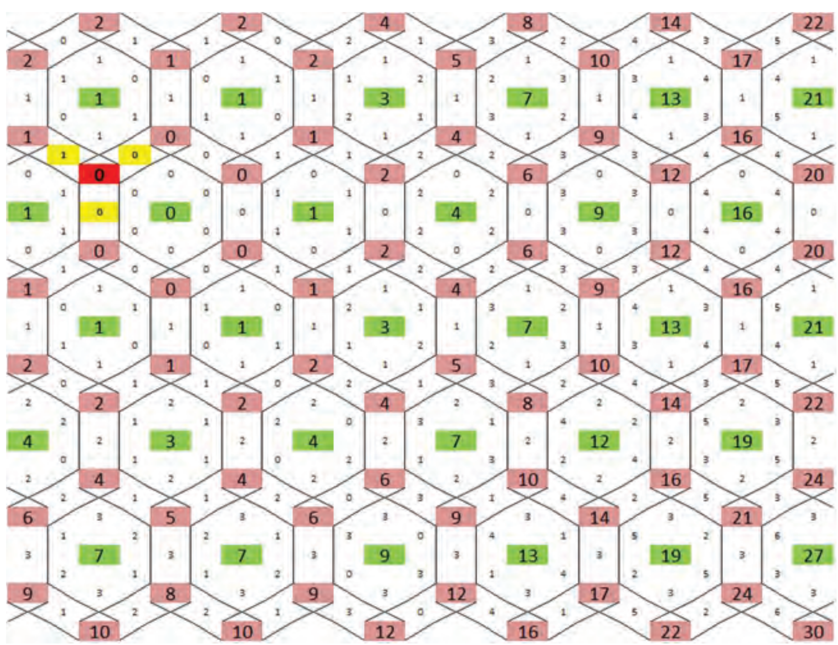


Figure 9: A “face first” projection.

3) Two slopes are zero and the other one is non-zero. This results in an entire hexagon touching the plane and the axis of symmetry of the parabolahedron being perpendicular to the plane through the center of that hexagon as in Figure 7.

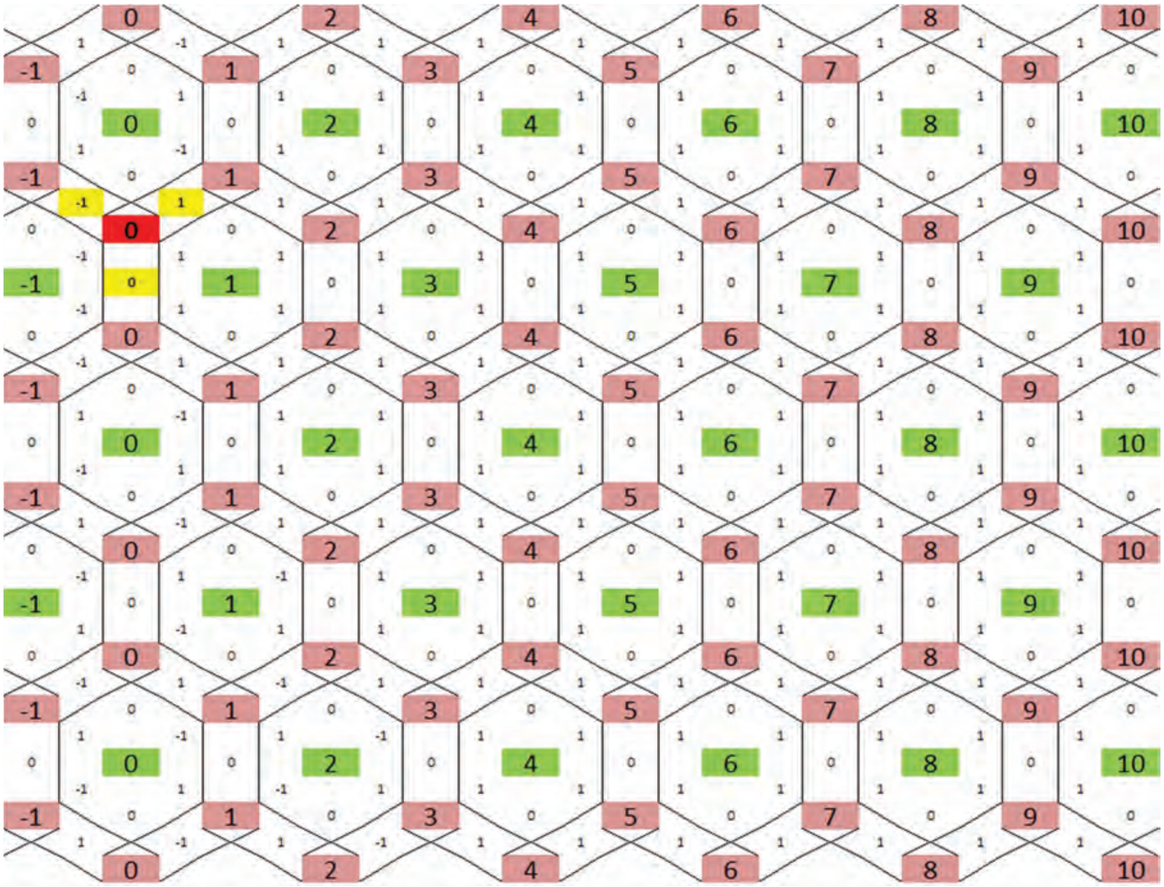


Figure 10: A plane is a degenerate parabolahedron.

4) One slopes is zero and the other two have the same absolute value, but with opposite signs. This results in a degenerate parabolahedron, in fact a plane, intersecting the original plane along the line extending from the origin and including the edge with 0 slope as in Figure 8.
5) Other variants exist including a variety of asymmetrical vertex first projections where the axis of symmetry is not through any of the vertices.
6) Using all negative slopes will generate an inverted parabolahedron with the bowl facing downward.
7) All zero slopes with a non-zero origin will result in a plane parallel to the original. This plane is another variant of a degenerate parabolahedron.

Further study:

The models generated by this algorithm can be used to feature many different aspects of mathematics and geometry.

In the simplest case, a predetermined grid like any of the examples shown can be the basis for stacking cereal or coins without even any focus on the resulting shape. Next in complexity might be the development of several different grid patterns like those featured here. Larger grids could be

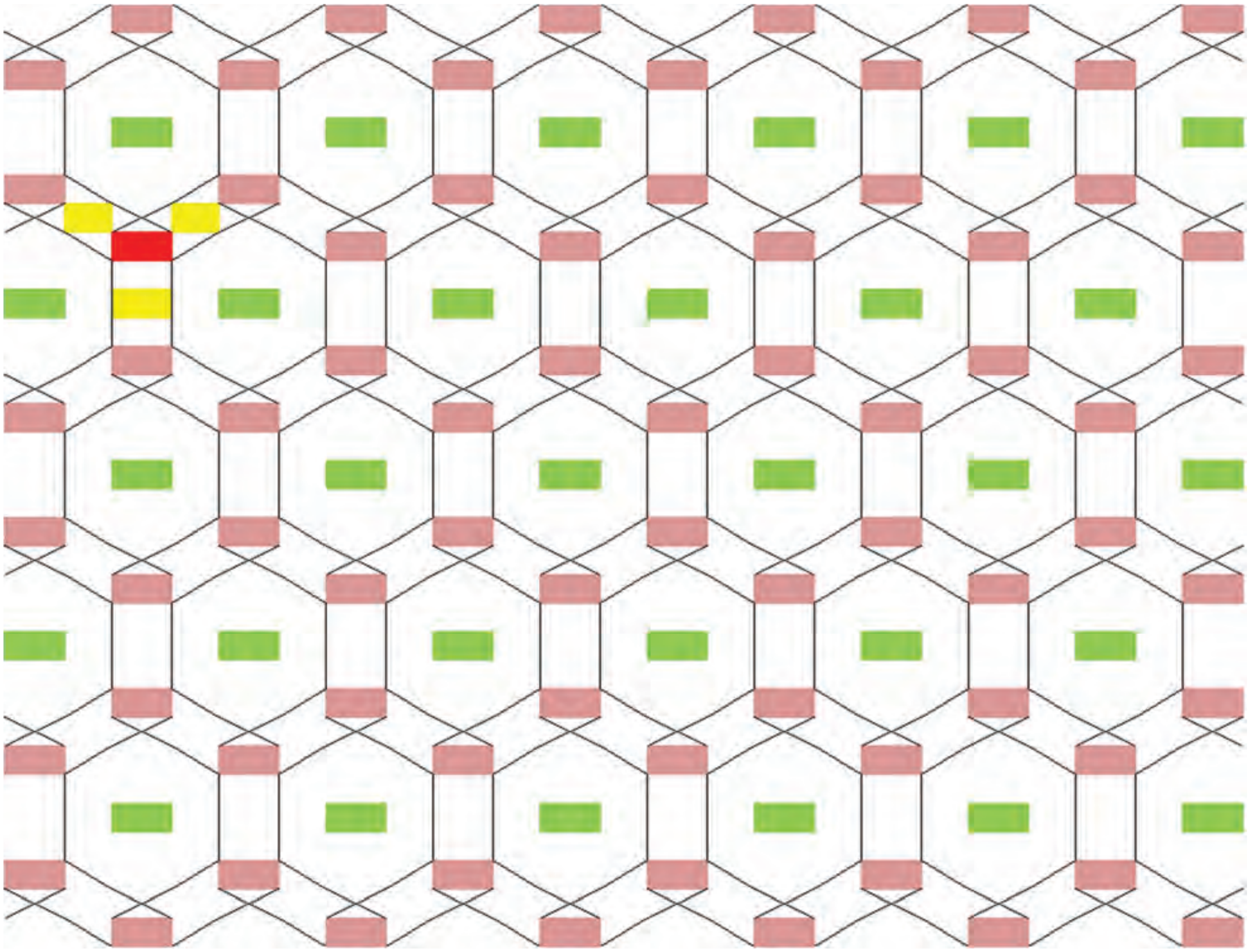
reproduced on posters allowing larger groups to collaborate on completing the cells. Further insights can be developed regarding rotational symmetries of the various patterns including 3-fold for the vertex first model, 2-fold or mirror symmetry for the edge first model, 6-fold symmetry for the face first model, as well as asymmetrical models. Intermediate studies can involve adding volumes of the hexagonal prisms to calculate a total volume of the parabolahedron. This is simpler than it may appear at first glance since the number in green represents the average height of all points on the facet, and therefore the volume of a given cell is simply the area of the hexagon multiplied by the number in green.

The model can also provide a centerpiece for more advanced level students to delve into discussions involving contour maps, calculating the slope of individual facets, especially those which don't lie on lines of symmetry. Other areas to be explored include exponential growth of the Z coordinates and vector addition of slope vectors in any given cell. Parallel slices of the parabolahedron can be discovered by observing concentric "rings" of hexagons around any chosen vertex or face, not just the origin.

Affine transformations and change of basis can be modeled by examining the cases where the three pairs of hexagon sides are different lengths, or the Z axis is not perpendicular to the grid. It may come as a surprise that the three cells surrounding the initial point are not necessarily coplanar as all our examples have been. These first three edges could be replaced by three non-coplanar axes such as the standard X, Y and Z axes which would result in an axis of symmetry which is not orthogonal to any of them. This model can even be used to generate a four-dimensional parabolahedron. Just drop me an email and ask me about it.

Enjoy and let me know what other insights you glean from this delightful geometric construct.

Honeycomb grid worksheet



Ring-a-ding Numeration

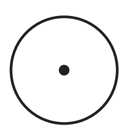
Jim Henle[1]

Ostensibly this paper is my gift to you, but really it's the website I created with animation that you generate. What you are reading now is a simple introduction.[2]

Ring-a-ding numeration is a simple system for writing numbers. It is, actually, a disguised version of binary. It's a pretty version, but it's totally impractical. It does have a lovely visual process for adding numbers, but it's singular feature is that it also has a visual process for multiplying numbers. I can't think of any other numeration system where multiplication is attractive.[3]

Here's the idea. A dot represents the number 1.

A ring doubles whatever is inside it.

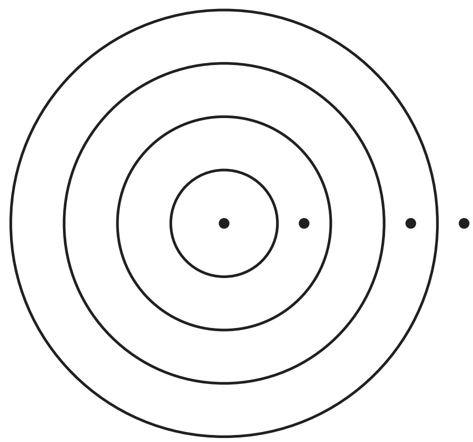


Here's the numeral for 27:

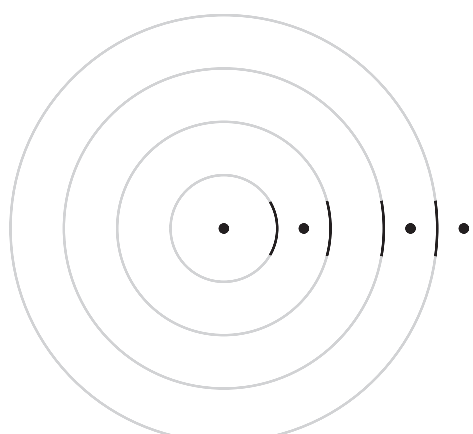
[1]With much thanks to Fred Henle for encouragement and technical assistance.

[2]And some of this is excerpted from a column I wrote for the *Mathematical Intelligencer*, "The Same, Only Different" 39(2): 60-63, 2017.

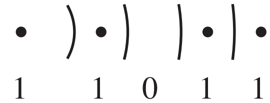
[3]Possible exception: Napier's bones?



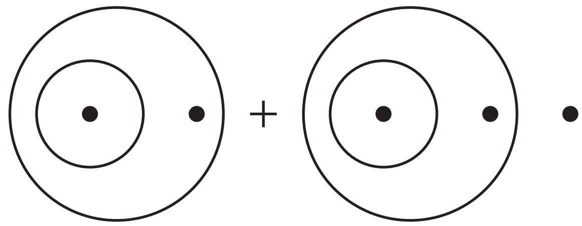
Going from outside in (right to left), the dots are worth 1, 2, 8, and 16. If you just look at a piece of this,



you can see the connection to binary:



To add numbers, you place them side-by-side,

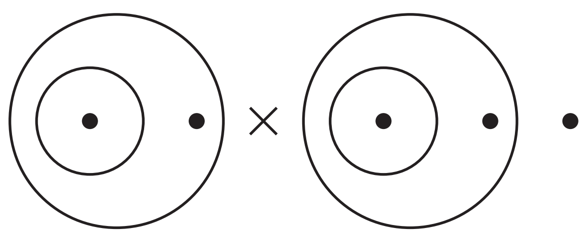


and smoosh them together. The smooshing process is fun, a little like cells splitting, but in reverse. You smoosh so that the numeral consists of nested circles with no more than one dot per region.

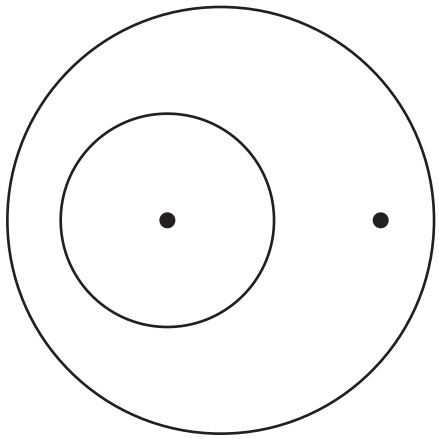
I could show you pictures, but the animated version (on the website) is more satisfying.

www.math.smith.edu/ jhenle/Ringading

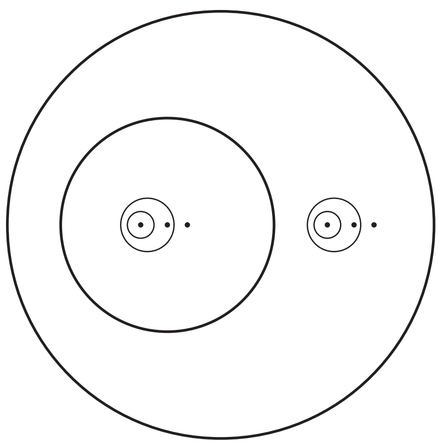
And to multiplying two numbers,



you simply replace every dot in one numeral



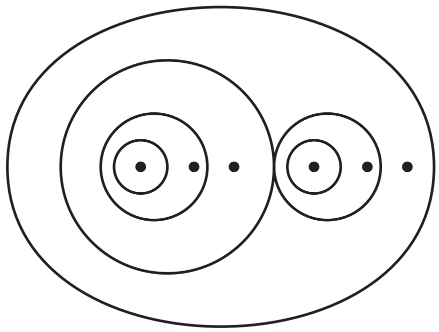
with the other numeral.



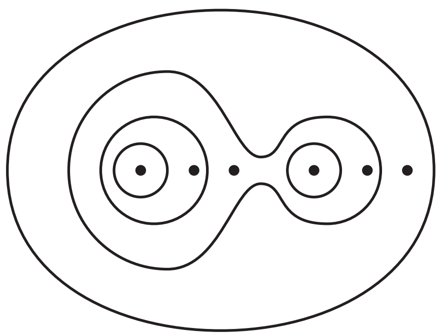
And then smoosh

But we’re not done yet. If we allowed pictures like this, then numbers wouldn’t be represented uniquely. To insure uniqueness, we insist that the curves are all nested. I also insist that no more than one dot appear in any region.

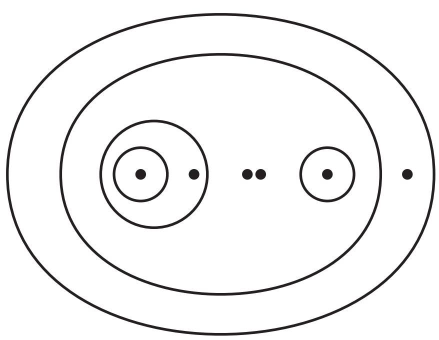
Fixing this numeral above goes like this: A circle on the left goes up to a circle on the right,



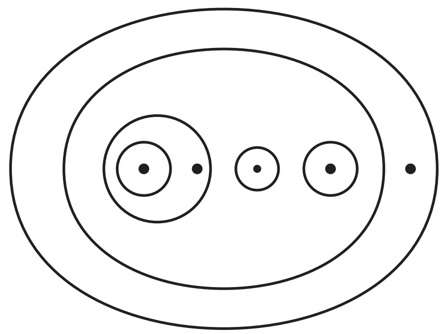
and they do something affectionate.



And now we have two dots in one region. So of course the dots get together.



It's love. They produce a baby and disappear.



To see this all happen in real time, go to

www.math.smith.edu/ jhenle/Ringading

LUCKY 13

by Kate Jones – presented at G4G13

What is 13? A famous digit,
A prime, a luck sign, that makes people fidget.

Today I bring you, to win your smiles,
13 sets of pretty tiles.
Such geometry and tessellations
Make for sweet math recreations.

These are the brainchild of Michael Dowle,
A British scientist always on the prowl
For designing puzzles with many solutions,
Each with its 13 convolutions.

Behold how 13 different tiles comport
When ins and outs a hexagon distort.
In their assembled star arrays
Each tile in turn the center spot displays.

Observe now how some tiles are chiral.
That's an idea that could go viral.
And some have rotation on their mind
While others reflect the mirror kind.

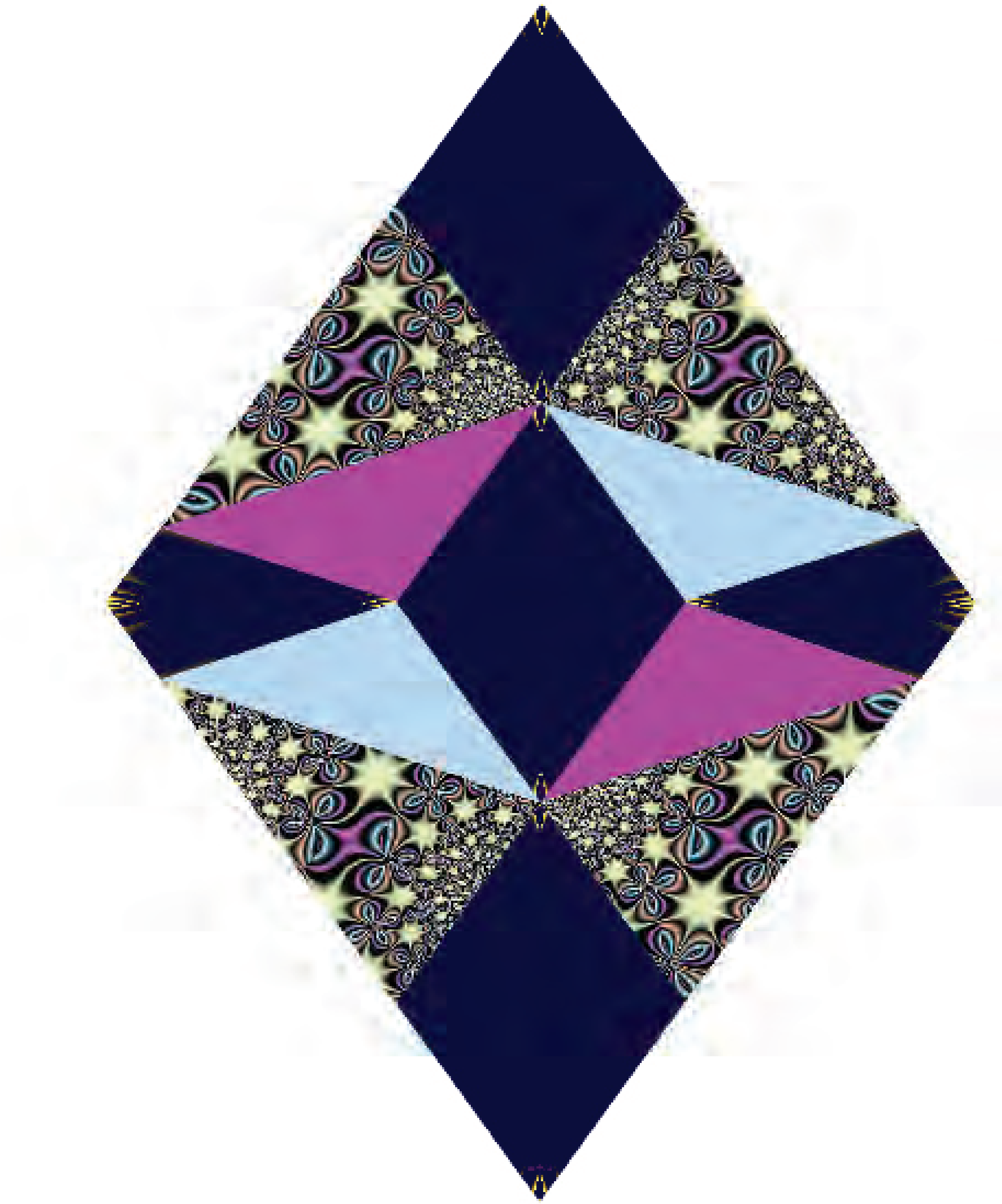
This last one, "Leaves", is our own production
From Jacques Griffioen's first introduction.
Its edge-wise connections and symmetry cycles
Differ in style and template from Michael's.

And here at G4G this year
I present the world premiere
Of one of Michael's lucky stars.
We hereby launch the "Cookie Jars".

Besides the star, a tasty trove of treasures
Await your pondering and solving pleasures.
It's your lucky day such play to enable –
Come see them at my sales room table!

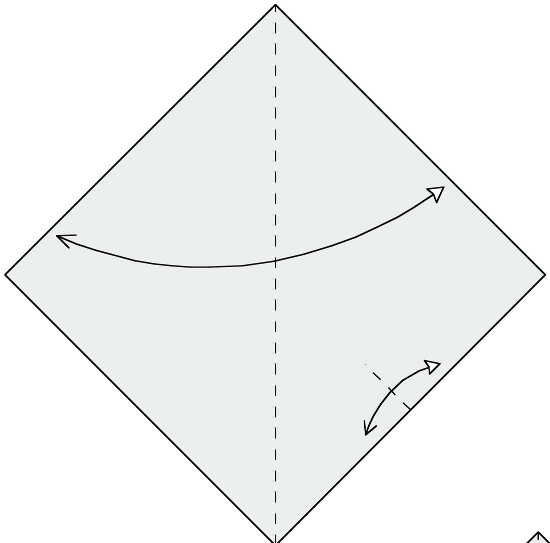
Golden Magic

Matjuska Teja Krasek

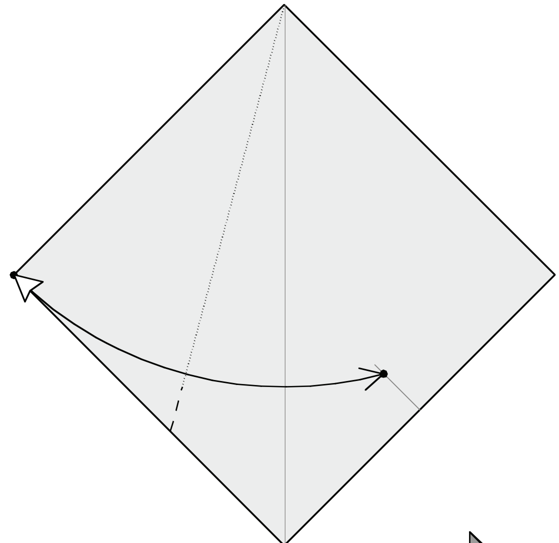


RiF-RiF Bird

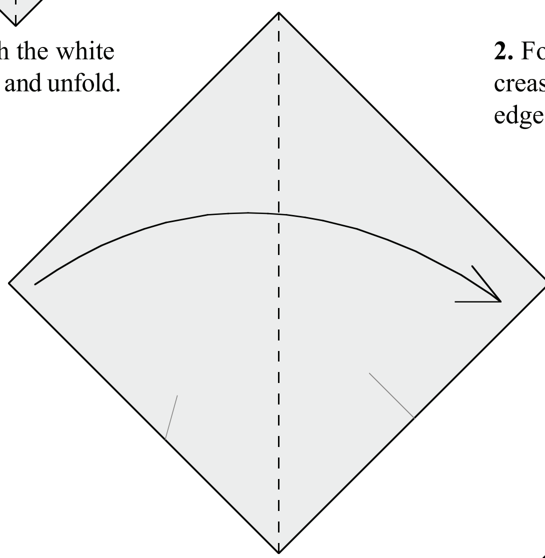
RiF-RiF stands for "Rigidly Foldable, Rigidly Flapping." This bird (which uses a mechanism derived from that of Randlett's "New Flapping Bird") can be folded rigidly and flaps rigidly, i.e., with all deformations happening on creases and no bending of facets.



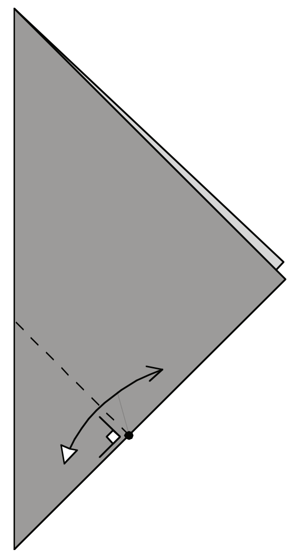
1. Begin with the white side up. Fold and unfold.



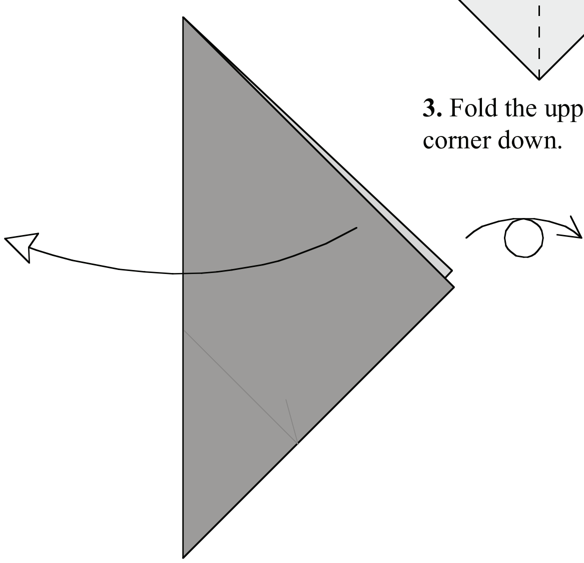
2. Fold the point to the crease, pinch along the edge, and unfold.



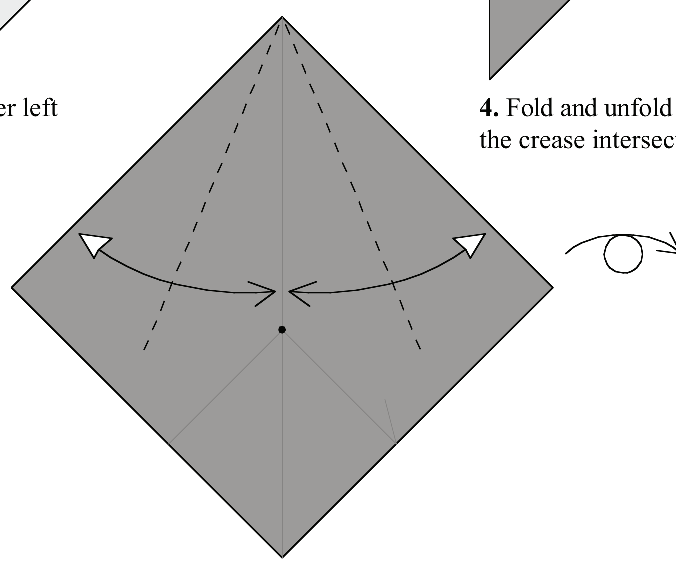
3. Fold the upper left corner down.



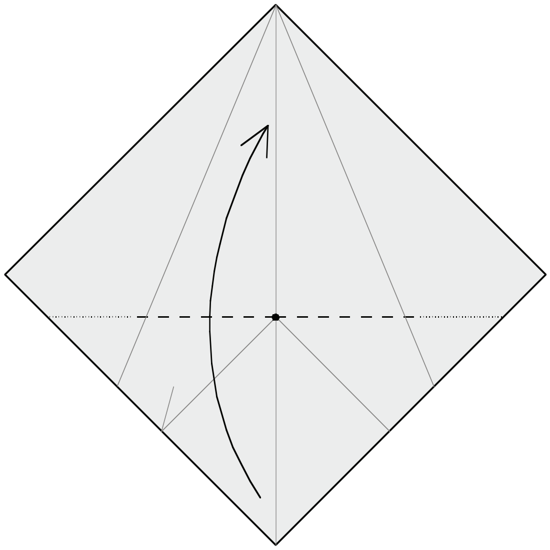
4. Fold and unfold through the crease intersection.



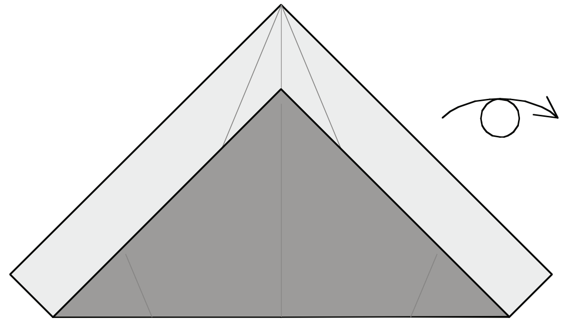
5. Unfold, then turn the paper over.



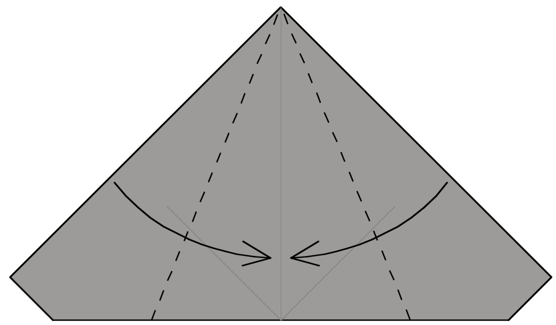
6. Fold the edges to the diagonal (they don't need to extend all the way across the paper) and unfold. Turn the paper back over.



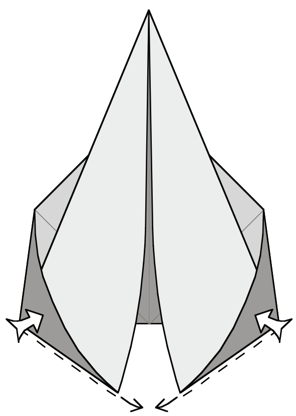
7. Fold the lower right corner up along a fold that passes through a crease intersection. You don't need to make it sharp beyond the creases you just made.



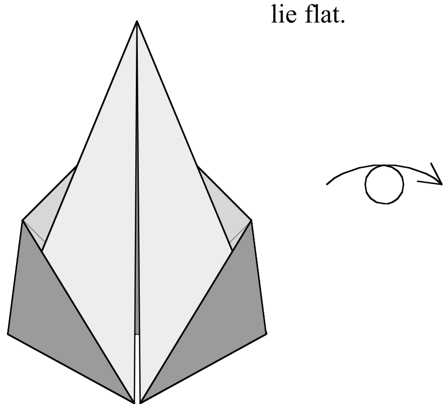
8. Turn the paper over.



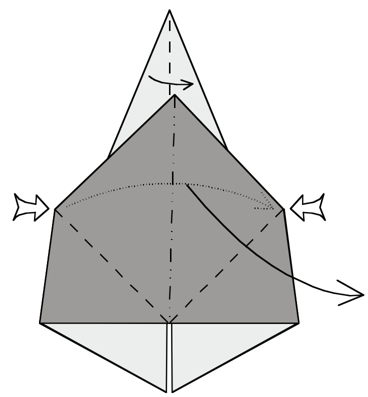
9. Fold the side edges in on the existing creases. The paper will not lie flat.



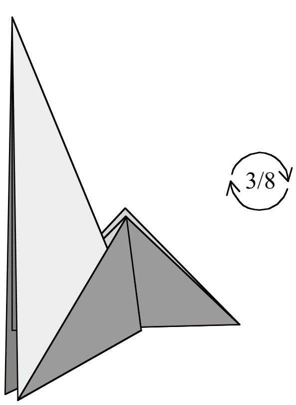
10. Flatten the paper, forming creases that run to the corners.



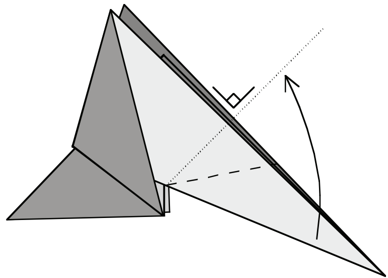
11. Like this. Turn the paper over.



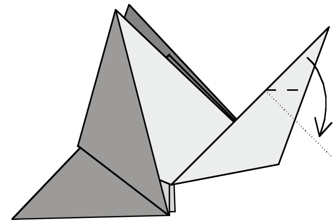
12. Squeeze the sides together and open out the pocket between the two top flaps. Flatten.



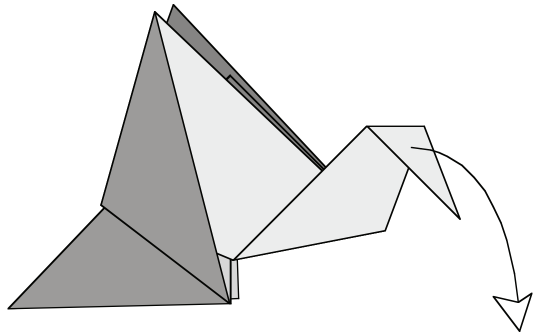
13. Rotate the paper 3/8 turn.



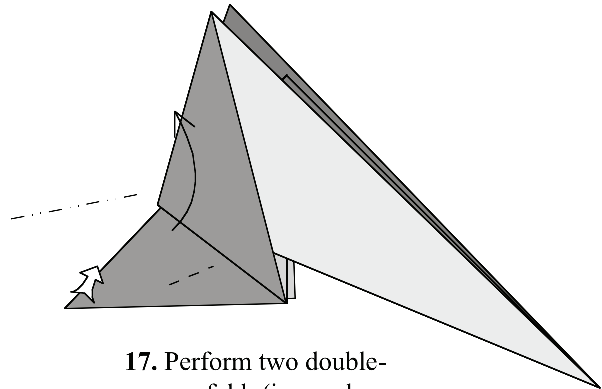
14. Fold the corner up, forming a right angle.



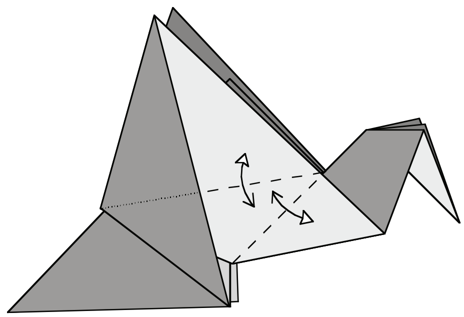
15. Fold the corner down to form a head.



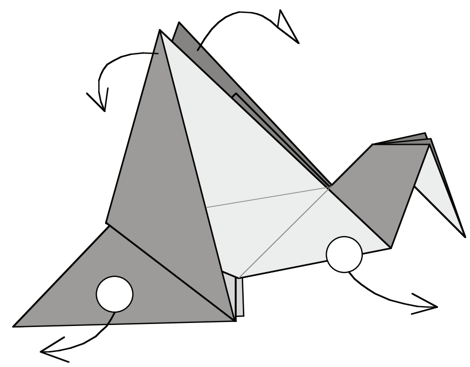
16. Unfold to step 14.



17. Perform two double-reverse-folds (i.e., each reverse fold actually consists of two side-by-side reverse folds).



18. Crease through all layers shown to form distinct hinges.



19. Hold where shown and pull; the bird will flap rigidly. Once the creases are in place, it can also be folded and unfolded rigidly in a single motion.

Klein Bagel Baby Toy

Sabetta Matsumoto
School of Physics
Georgia Institute of Technology

Immersions of Klein bottles have taken hold in the popular imagination. This single sided surface appears in sculpture, jewelry, graphic art, clothing, accessories, and more. The Banchoff-Cervone bottle, shaped like an elegant crystal decanter, is the ubiquitous immersion. This pattern, however, features the elegant symmetry of the Klein Bagel immersion. The Klein Bagel immersion demonstrates that the Klein bottle is made by gluing two Möbius strips together along the central branch line.

This pattern is knit in a figure-eight shape, using stockinette on one lobe and reverse stockinette on the other. The unusual technique in the cast-on used to achieve this enables the non-orientability of the Klein bottle to show through. This pattern uses seven colours of yarn, but it can be adapted to accommodate any odd number of colours and maintain the same notion of non-orientability

Finished Measurements
Diameter: 7 inches.

Materials
7 skeins KnitPicks Comfy Worsted [75% Pima Cotton, 25% Acrylic; 109yd per 50 gram ball]; Color: Zinnia, 1 ball; Carrot, 1 ball; Semolina, 1 ball; Peapod, 1 ball; Marina, 1 ball; Marlin, 1 ball, Lilac, 1 ball
2 16-inch US #5/3.5mm circular needles
1 set US #5/3.5mm double pointed needles
stitch markers
tapestry needle
crochet hook

Gauge
20 sts and 28 rows = 4 inches in stockinette stitch.

Directions

Using waste yarn CO 40 sts using a crochet cast on. Place markers between the 10th and 11th sts and the 30th and 31sts. Equally divide the sts between two needles (Fig. 2a), being careful not to twist the yarn. Cross the second needle over the first to form a figure eight (Fig. 2b). The stitch markers should meet at the intersection.

Be careful not to twist your piece at any time. It may take a few times for the piece to properly get started. You should know by the first 3-4 rows if your needles are properly crossing over one another at the centre of the figure eight.

Row 1a: in Peapod. Starting with the side of the figure-eight closest to you, sl 10. Your needle with be below the cable of the other pair of needles. Slide the needle out from this and holding the life yarn on top of the work, K 10. These 10 stitches will be on the inside of the left lobe of the figure-eight.
Row 1b: Rotating the piece clockwise from the top, switch to the other needles. K10. These 10 stitches will be on the inside back of the right lobe. When you come to the middle you will find your left needle below the cable of the

((a)) Row 1a.

((b)) Row 1b.

((c)) Row 2.

((d)) Row 25.

other needle and the last row of yarn on the previously worked side. Pull your left needle tip out from under both of these and continue working the stitches with your live yarn on top. The next 10 stitches will be on the outside front of the left lobe of the figure eight.
Row 1c: Rotating the piece clockwise from the top, switch to the other needles. K10. These 10 stitches will be on the outside front of the right lobe. When you come to the middle you will find your left needle below the cable of the other needle and the last row of yarn on the previously worked side. Pull your left needle tip out from under both of these and continue working the stitches in pattern with your live yarn on top. The next 10 stitches will be on the inside back of the left lobe of the figure eight.
Rows 2-25: Repeat rows 1b-1c 24 times. *Row 26a:* Repeat row 1b. *Row 26b:* Rotating the piece clockwise from the top, switch to the other needles. K10. These 10 stitches will be on the outside front of the right lobe. When you come to the middle you will find your left needle below the cable of the other needle and the last row of yarn on the previously worked side. Pull your left needle tip out from under both of these. You will now switch colours. (Your piece should measure about 2.75" long.)

Row 27a: in Zinnia. K 10. These 10 stitches will be on the inside of the left lobe of the figure-eight.
Row 27b: Rotating the piece clockwise from the top, switch to the other needles. P10. These 10 stitches will be on the inside back of the right lobe. When you come to the middle you will find your left needle below the cable of the other needle and the last row of yarn on the previously worked side. Pull your left needle tip out from under both of these and continue working the stitches with your live yarn on top. The next 10 stitches will be on the outside front of the left lobe of the figure eight.
Row 27c: Rotating the piece clockwise from the top, switch to the other needles. P10. These 10 stitches will be on the outside front of the right lobe. When you come to the middle you will find your left needle below the cable of the other needle and the last row of yarn on the previously worked side. Pull your left needle tip out from under both of these and continue working the stitches in pattern with your live yarn on top. The next 10 stitches will be on the inside back of the left lobe of the figure eight.
Rows 28-51: Repeat rows 27b-27c 24 times.
Row 52a: Repeat row 27b.
Row 52b: Rotating the piece clockwise from the top, switch to the other needles. P10. These 10 stitches will be on the outside front of the right lobe. When you come to the middle you will find your left needle below the cable of the other needle and the last row of yarn on the previously worked side. Pull your left needle tip out from under both of these. You will now switch colours. (Your piece should measure about 5.5" long.)

Rows 53-78: in Marlin Repeat rows 1-26.

Rows 79-104: in Semolina Repeat rows 27-52.

Rows 105-130: in Marina Repeat rows 1-26.

Rows 131-156: in Lilac Repeat rows 27-52.

Rows 105-180: in Carrot Repeat rows 1-24.
Row 181a: Repeat row 1b.
Row 181b: Repeat row 26b.

Finishing
Stuff both tubes. Remove the provisional cast on and places stitches on dpns. Add a π-twist to the piece, and use the remaining Carrot yarn to Kitchener both tubes shut in pattern. Sew in the ends.

The Eternal Magic of The Shape

- A poetic gesticulation

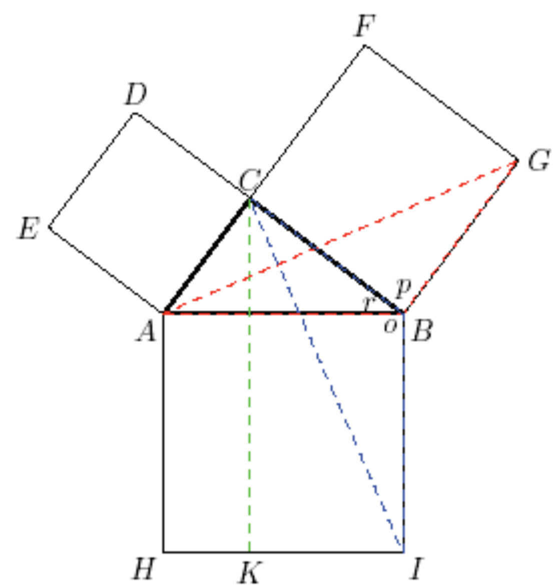
Hans Christian Andersen wrote the Danish poem on how to solve Pythagoras' Theorem in 1831, "Formens Evige Magie")

Translated by Jane Møller Nash, janemnash@gmail.com

Whether the baking tray, or the cake itself
is the main thing
in this world, we will leave out.
I bring - (yes, we will end up with the same result)
I bring a small frame
for which I wrote so-called poetry.
And perhaps the frame has the most value,
because it has "the shape's eternal magic"
and it can beat the heart's poetry.
He who, to date, rejected every piece
I brought forth (because there was a shadow therein),
maybe my frame brings happiness to him,
because I will put it into the shape;
I am going to pluck this prose-heather,
and, in short - make soup on a stick.
What is to poetry most contradictory,
Geometry's favorite Master
Matheseos, here on the sheet I write;
Now look! Watch out, everyone.

The triangle *ABC* is given here,
right angled and on its sides reside squares.
The proof is now whether the two areas,
i.e. the two squares on each short side
AC, *BC* (I name these),
have just that same size as the area
of the square on the hypotenuse.
Now, let us go to our preparations.

A vertical line, as you know, must be
drawn to the longer side,
and then extended to *K*.
Then you will find, not the least is missing,
The *AB* square is
split (as *AK*, *BK*) into two rectangles.
(Because, as you know, two lines
share the property,
when they vertically stand on a third line,
they are in fact parallel.)
Now draw lines from *A* to *G*, from *C* to *I*,
and now the preparation is over.
Not true, oh Master! – don't threaten, not
with the whisk!



Now, let us go to the proof.
- We have the two triangles, *ABG*
and *CBI*, here the angle *p*
is equal the angle *o*, but *o* is a right angle,
Yes, there is no one who will deny that,
for right angles are in squares.
Now the angle *r* is equal to the angle r. Right?
(It is common sense that
a size is equal to itself.)
Thus, *p* plus *r* is equal to *o* plus *r*
(here in the figure, the two smaller squares).
Now an equal amount is added to both
creating two larger sums.
(Now, the proof soon is over,

It goes towards the end.)
Look, the angle *ABG* is equal to *CBI*,
AB is equal to *BI*, *BG* is equal to *BC*
(in a square all sides are equal,
therefore, as true as three is always three,
two sides and an angle will help us).
We dare to set the triangle *ABG*
equal to *CBI* (and that is no coincidence).
Now *ABG* is equal to half a *BF*,
beware!
Now *CBI* is half a *BK*.
(Remember: equal sizes are the same.)
Equal are the denominators, equal are the numerator;
and equal will the quotient be,
and by this we get:
AD is equal to *AK*.

So here you have the method,
soon as Pythagoras you will solve the riddle.

Yes, solved, proved - the great sorcery!
Heavenly thanks! - It's over!
For such verses are not trickery;
they run well as if there was nothing within -
however, in here was common sense and shape-magic.
(Last, I hope,
and at least this shape is free
from what badly dampen every melody:
a drop of mud.)
Reason and shape have created poetry.
Here you can see "The Eternal Magic of Tthe Shape"

H.C.Andersen's original poem in nearly 200 years old Danish

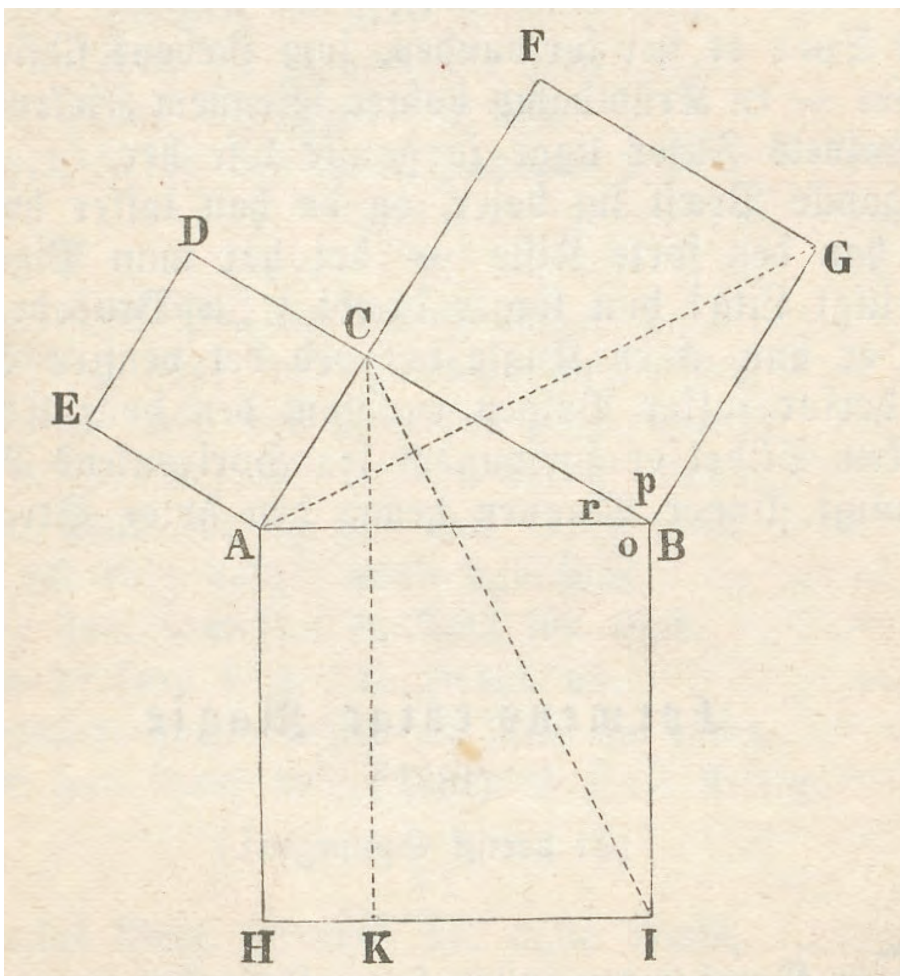
Om kageformen, eller selve kagen
er hovedsagen
i denne verden, går vi her forbi.
Jeg bringer— (ja, det kommer til det samme)
jeg bringer nemlig her en lille ramme
til hvad jeg skrev og kaldte poesi.
Og muligvis får rammen mest værdi,
thi den har „formens evige magi"
og den kan stikke hjertets poesi.
Han, som til dato vragede hvert stykke,
jeg bragte frem (fordi deri var skygge),
måske hos ham min ramme gør sin lykke,
thi jeg skal trænge den i formen ind;
jeg vil den seje prosa-lyng oprykke,
og, kort sagt—lave suppe på en pind.
Hvad der er mest mod poesien bister,
geometriens yndede magister
Matheseos, jeg her på bladet rister;
se så! pas på enhver.

Trianglen ABC er givet her,
retvinklet og på siderne kvadrater;
beviset er nu om de to krabater,
det, at kvadraterne på hvert kateder
AC, BC (jeg nævne disse steder)
er just i et og alt, som den krabat,
hypotenusen kalder sit kvadrat.
Nu går vi da til vore præparater.
En lodret linie må man som De ved

her drage til den større side ned,
og så forlænge den endnu til K,
da vil man finde, ej det mindste mangler,
AB-kvadratet ganske rigtigt stå
delt (som AK, BK) i to rektangler.
(Thi tvende linier, man ved,
har just det generelle,
når på en tredie de stå lodret ned,
så er de også ganske parallelle.)
Nu drages en fra A til G, fra C til I,
og da præparationen er forbi.
Ej sandt, o mester! —true dog ej med riset!
Nu går vi til beviset.
— Vi har de to triangler ABG
og CBI, hos dem er vinklen p
lig vinklen o, men o er lig en ret,
ja, der er ingen, som vil nægte det,
thi rette vinkler er der i kvadrater.
Nu vinklen r lig vinklen r. Ej sandt?
(Thi sund fornuft kan sige
hver størrelse jo med sig selv er lige.)
Således p plus r lig o plus r man fandt
(her i figuren står de små krabater).
Når lige nu til begge bliver lagt,
en lige sum er da tilvejebragt.
(Nu er vi med beviset snart forbi,

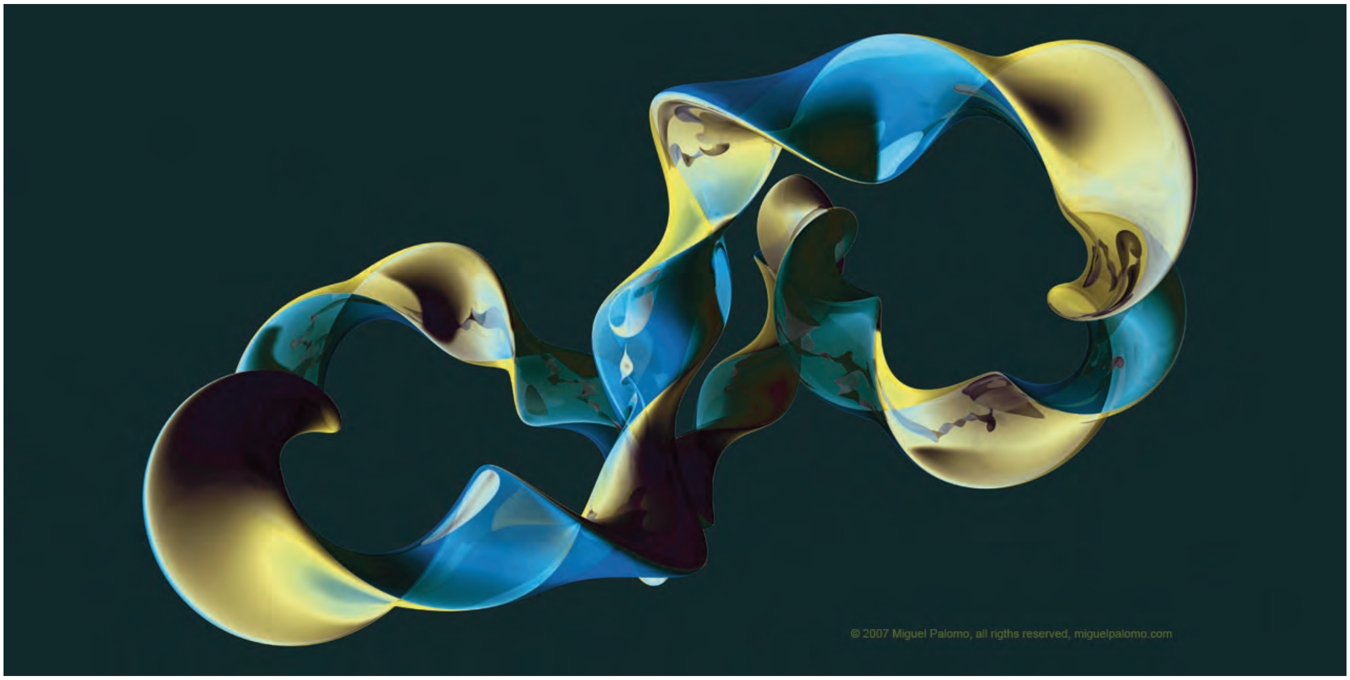
det stærkt mod enden lider.)
Se vinklen ABG lig CBI,
AB er lig BI, BG er lig BC
(i et kvadrat er lige store sider,
derfor, så sandt som tre gør altid tre,
to sider og en vinkel vil os lette),
trianglen ABG vi her tør sætte
lig CBI (og det er intet træf).
Nu ABG er lig en halv BF,
pas på!
Nu CBI er lig en halv BK.
(Husk: lige stort for lige stort kan gå.)
Ens er divisor, ens er dividenden;
ens bliver altså også kvotienten,
og ad den samme vej vi få:
AD er lig AK.
Der har du måden,
snart som Pythagoras man løser gåden.
Ja løst, bevist—du store trylleri!

Du himmel tak! —at det er nu forbi!
Thi slige vers er ikke narreri;
de løbe vel, som der var intet i—
dog her var jo fornuft og form-magi.
(Det sidste vil jeg håbe,
og denne form er i det mindste fri
for hvad der dæmper slemt hver melodi:
en mudderdråbe.)
Fornuft og form har her skabt—poesi.
Her ser man „formens evige magi."

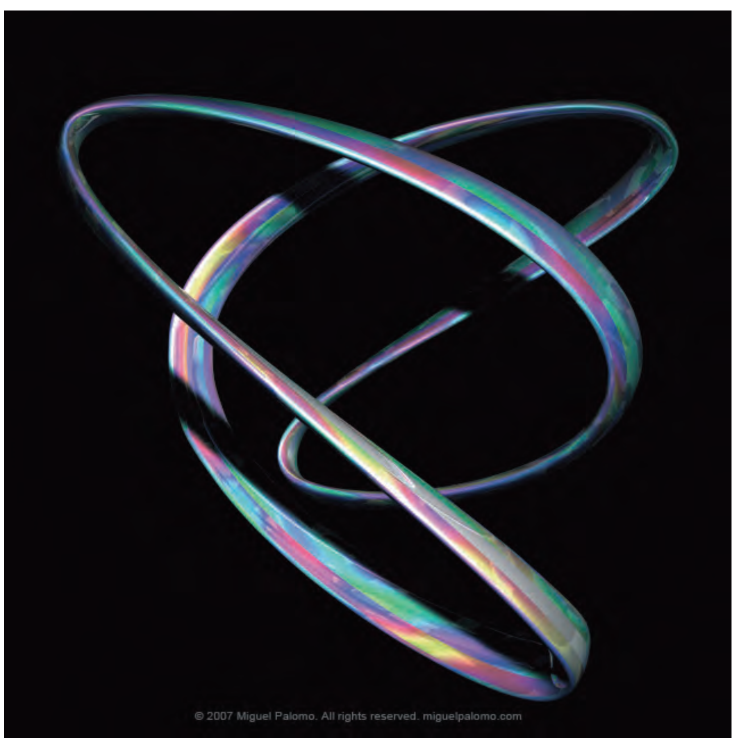


Math Art

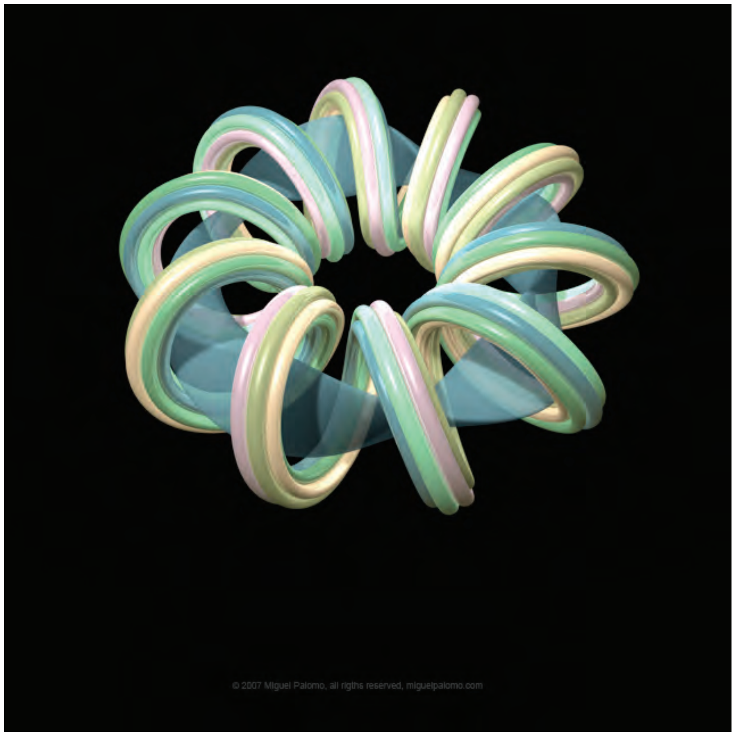
Miguel Palomo



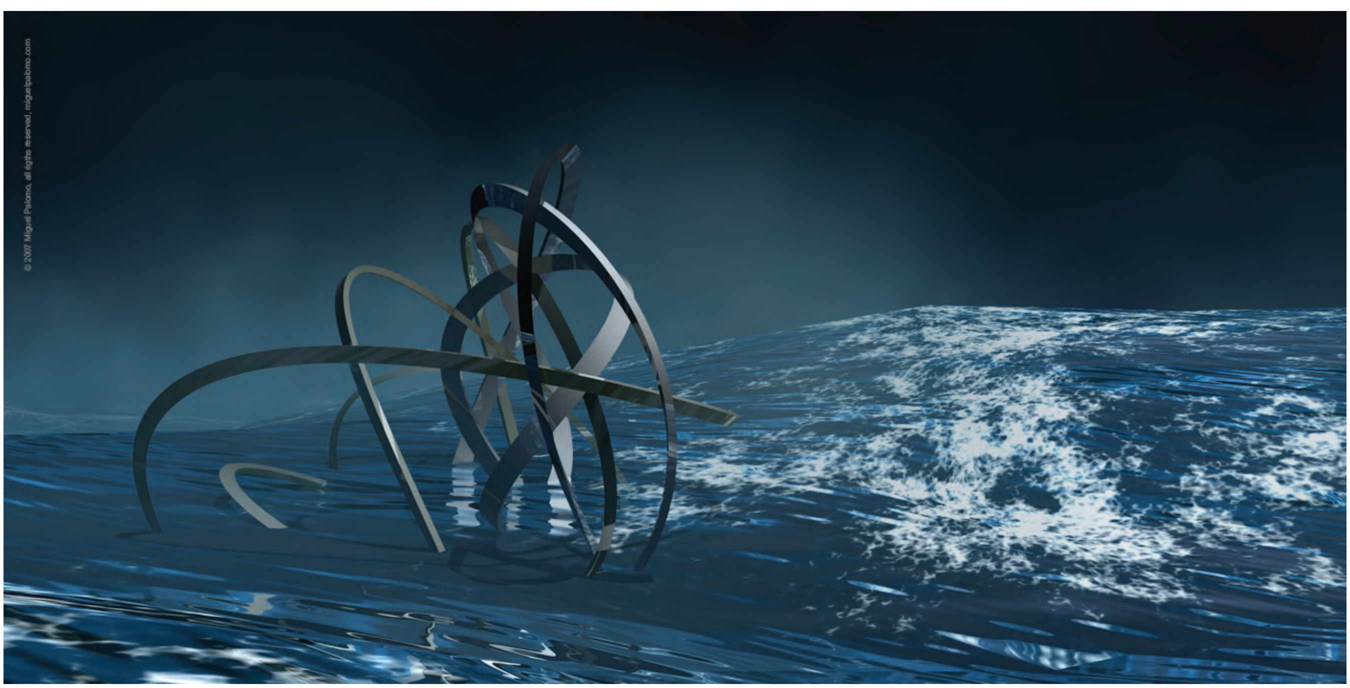
There Are Eight Things in You that Thrill Me



I Have Always Dreamt About You



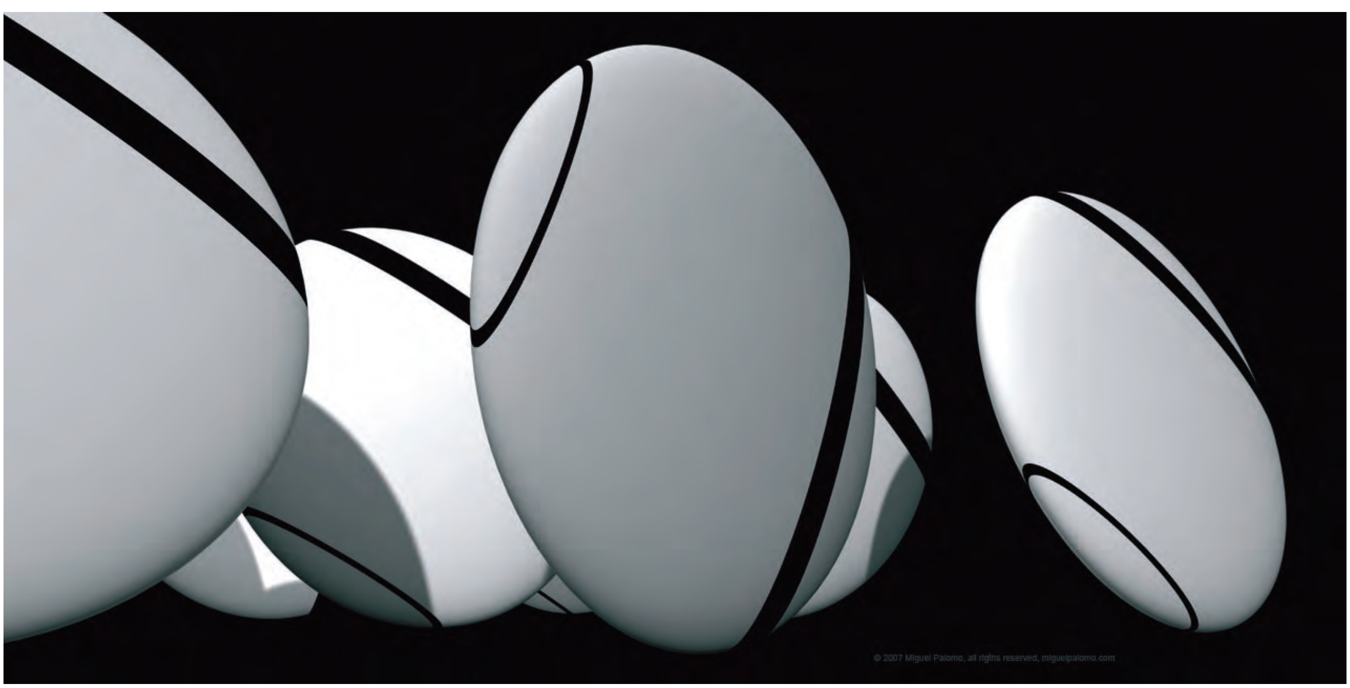
Gentle Hold



Metal Splendor

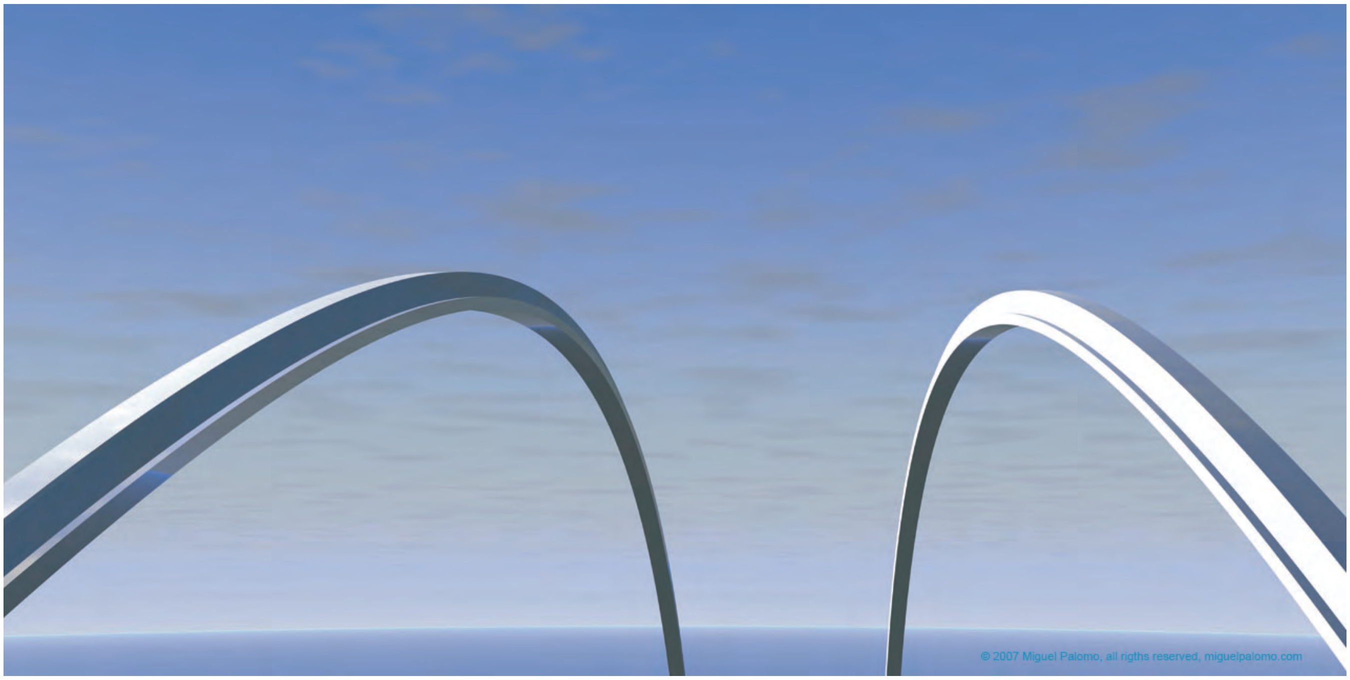
I Let Water Purify Me

Pure Jewel



The Way of Love

There is Something Amazing in All This



Almost There

GATHERING 4 GARDNER 13

MATHS EVERYWHERE

http://music.miguelpalomo.com

It’s a mystery that induces awe
It’s a vision of another world
All the concepts all the theorems
All the numbers there and everywhere
All the symbols all the truths
In equations of eternal youth
Out of silence out of empty space
Zero rises in a daze

Maths everywhere
Subtle poetry of numbers
Logic, relations and shapes
Maths everywhere
In the eye of the beholder
Beyond the mind’s greatest depths

Away away from the particular
Lies a wider view that of Algebra
Sheer beauty in its purest form
Operations glowing at the core
Elevation with a gentle push
Formal action in a quest for truth
Master Euler be my guide
Live forever in my mind

Maths everywhere...

In a feeling there is shape
Like there is in everything else
Geometrical intuition
King of mathematical tradition
Little truths little things
Shining brighter than unproven myths
Master Euclid here I am
Help me out to prove the facts

Maths everywhere...

I Heart Cardioids

Make a Cardioid Flip Book

David Richeson

Roll a circle around another circle of the same radius. A marked point on the first circle traces a curve called a *cardioid*. (In figure 1 we rolled the orange circle around the red circle to draw the green cardioid.) This beautiful heart-shaped curve shows up in some of the most unexpected places. Grab a cup of coffee and we'll show you some.

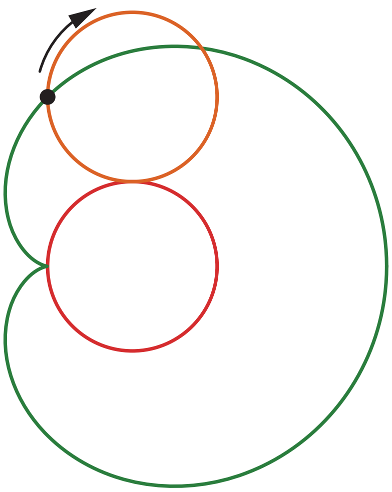


Figure 1. Roll a circle around another circle of the same radius and a point on the first circle traces a cardioid.

We do not know who discovered the cardioid. In 1637 Étienne Pascal—Blaise's father—introduced the relative of the cardioid, the limaçon, but not the cardioid itself. Seven decades later, in 1708, Philippe de la Hire computed the length of the cardioid—so perhaps he discovered it. In 1741, Johann Castillon gave the cardioid its name.

Got your coffee? Turn on the flashlight feature of your phone and shine the light into the cup from the side. The light reflects off the sides of the cup and forms a caustic on the surface of the coffee (see figure 2). This caustic is a cardioid.

The Mandelbrot set is one of the most beautiful images in all of mathematics (see figure 3). It is the set of complex numbers c such that the number 0 does not diverge to infinity under repeated iterations of the function $f_c(z) = z^2 + c$. The Mandelbrot set consists of a heart-shaped region with infinitely many circles, spiny antennae, and other heart-shaped regions growing off of it. That main heart-shaped region? It's a cardioid.

Figure 2. Shine a light from the edge of a coffee cup and the caustic it forms is a cardioid.

Cardioids even show up in audio engineering. Sometimes engineers need a uni-directional microphone—one that is very sensitive to sounds directly in front of the microphone and less sensitive to sounds next to or behind it. When they do, they reach for a *cardioid microphone*. The microphone is so-named because the graph of the sensitivity of the microphone in polar coordinates is a cardioid.

In this article, we present a few favorite places that cardioids appear. In particular, we will look at how we can use lines to construct the curved cardioid. At the end of the article, we provide a template that you can use to make your own cardioid. And we provided printable pages that can be used to make a cardioid flip book.

The Envelope of a Family of Curves

A common kids math doodle is to draw a set of coordinate axes and then draw line segments from $(0, 10)$ to $(1, 0)$, from $(0, 9)$ to $(2, 0)$, and so on, as in figure 4. This procedure magically produces a suite of lines that, when viewed together, has what appears to be a curved boundary. This curve is called the *envelope* of the family

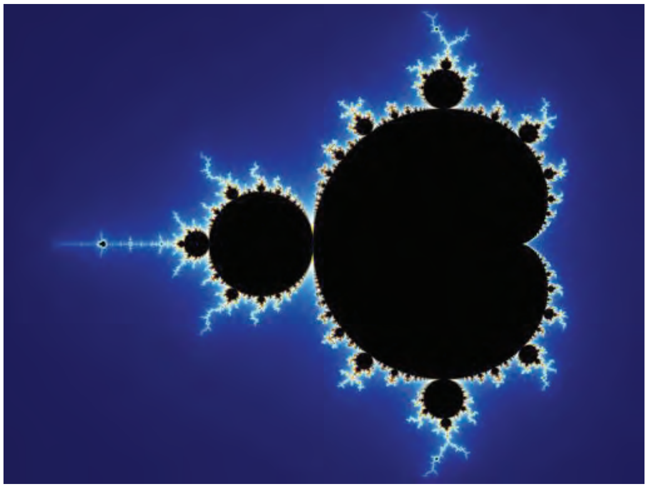


Figure 3. The main bulb of the Mandelbrot set is a cardioid.

of lines.

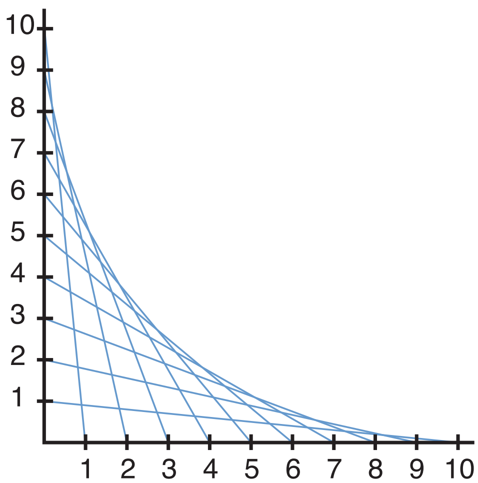


Figure 4. A curve as an envelope of lines.

Let C_t denote a family of curves parametrized by t. We can represent them as $F(x, y, t) = 0$ for some function $F : \mathbb{R}^3 \to \mathbb{R}$. For instance, in this elementary example, the line C_t joins $(0, 11-t)$ to $(t, 0)$, so it corresponds to $F(x, y, t) = yt + (11 - t)(x - t) = 0$.

Let us look at some features of this envelope. First, each line C_t is tangent to the curve. Second, if we take two nearby lines C_t and C_{t+h}, their point of intersection is near the curve, and taking the limit as $h \to 0$ yields a point on the curve. We could use either of these observations to produce a definition of an envelope, but instead, we use calculus.

In the following definition we let $F_t = \frac{\partial F}{\partial t}$ denote the partial derivative of F with respect to t.

Definition. Let $F : \mathbb{R}^3 \to \mathbb{R}$ be a differentiable function. The *envelope* of the set of curves $F(x, y, t) = 0$ is the set of points (x, y) such that both $F(x, y, t) = 0$ and $F_t(x, y, t) = 0$ for some value of t.

This is a mysterious definition. Why does it produce the envelope? For a fixed t and any $h \approx 0$, the curves $F(x, y, t) = 0$ and $F(x, y, t + h) = 0$ (that is, C_t and C_{t+h}) cross at a point near the envelope. Solving this pair of equations for x and y is equivalent to solving $F(x, y, t) = 0$ and $\frac{1}{h}(F(x, y, t + h) - F(x, y, t))$ for x and y. Then, as $h \to 0$, the point of intersection approaches a point on the curve. Thus, we find the point by solving $F(x, y, t) = 0$ and

$$\lim_{h \to 0} \frac{F(x, y, t + h) - F(x, y, t)}{h} = F_t(x, y, t) = 0$$

for x and y.

Returning to our example in figure 4, $F_t(x, y, t) = y - x - 11 + 2t$. If we set this expression equal to 0, solve for t, and substitute it into $F(x, y, t) = 0$, we obtain the equation $(x + y - 11)^2 - 4xy = 0$, which is a parabola opening along the line $y = x$. We can see this curve more clearly if we extend our figure beyond 1 through 10 (see figure 5).

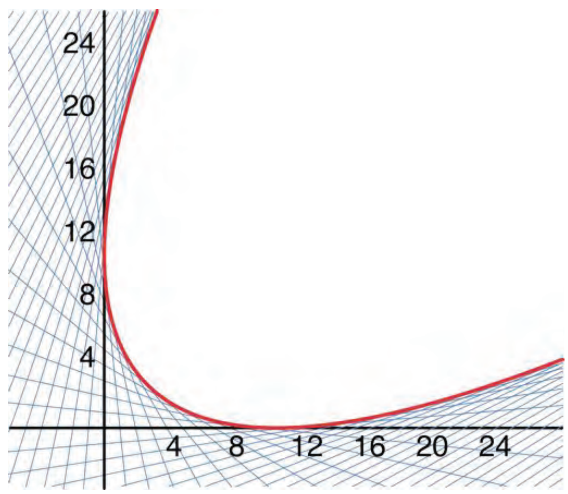


Figure 5. The envelope of lines is a parabola.

A Cardioid as an Envelope of Lines

It turns out that we can construct the cardioid as the envelope of curves, and we can do so in a number of different ways. For instance, pick a point P on a circle (the blue circle in figure 6, say). Draw circles with centers on the original circle that pass through P. The envelope of these circles is a cardioid.

But we will focus on a different example. Begin with a circle (the red circle in figure 7). Mark a certain number of evenly spaced points around the circle, N, say, and number them consecutively starting at some point P: $0, 1, 2, \ldots, N - 1$. Then for each n, draw a line between points n and $2n \pmod N$. In our example, $N = 54$, so we would join points 5 and 10, 19 and 38, and 31 and 8

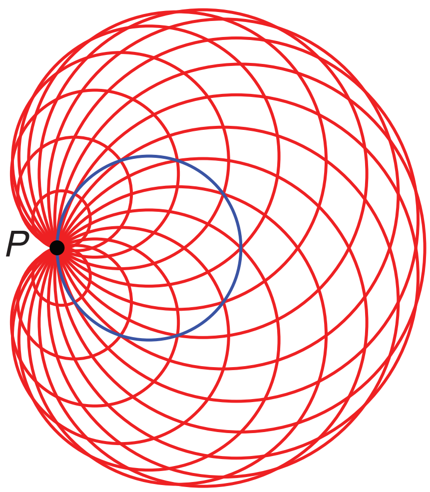


Figure 6. A cardioid as an envelope of circles.

(since 8 is 62 mod 54). The envelope of these lines is a cardioid.

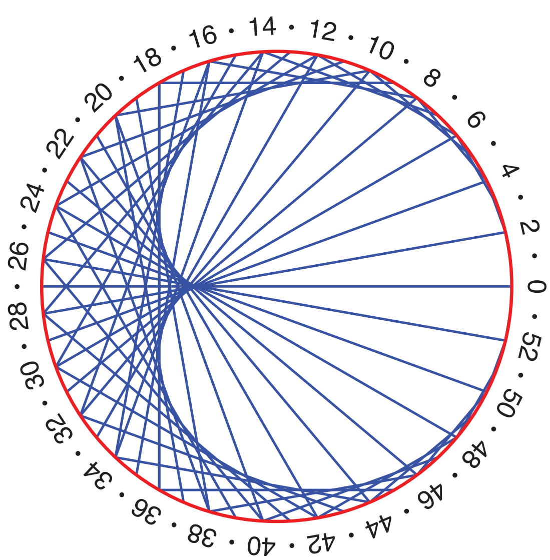


Figure 7. A cardioid as an envelope of lines.

Let's see why this is the case. Suppose our circle has center $(1,0)$ and radius 3 and that $P = (4,0)$. Now, starting at P, find points t and $2t$ radians around the circle from P, and draw the line segment joining them. We will show show that the envelope of all such lines is the cardioid with polar equation $r = 2(1+\cos\theta)$.

The two points on the circle—corresponding to t and $2t$—have coordinates $(3\cos t + 1, 3\sin t)$ and $(3\cos(2t) + 1, 3\sin(2t))$. The line joining them is

$$y - 3\sin t = \left(\frac{\sin(2t) - \sin t}{\cos(2t) - \cos t}\right)(x - 3\cos t - 1).$$

After some some algebra and some applications of double angle formulas, we can express this line as

$$(\cos(2t) - \cos t)y - (\sin(2t) - \sin t)x + \sin(2t) + 2\sin t = 0.$$

In particular, the expression on the left is our function $F(x,y,t)$. Taking the partial derivative of F with respect to t we obtain

$$\begin{aligned} F_t(x,y,t) = &(-2\sin(2t) + \sin t)y - (2\cos(2t) + \cos t)x \\ &+ 2\cos(2t) + 2\cos t. \end{aligned}$$

Now, we want to show that the x and y coordinates at which $F(x,y,t) = F_t(x,y,t) = 0$ is a point on the cardioid $r = 2(1+\cos\theta)$. The cardioid has one more surprise for us: This happens when $t = \theta$ (see figure 8)! We can express this polar curve with parametric equations as

$$\begin{aligned} x &= 2(1+\cos\theta)\cos\theta \\ y &= 2(1+\cos\theta)\sin\theta. \end{aligned}$$

And when we replace θ with t and substitute these expressions for x and y in F and F_t, we obtain 0. (The tedious calculations require both algebra and further applications of the double angle formula.) Thus, the cardioid is the envelope of this family of lines.

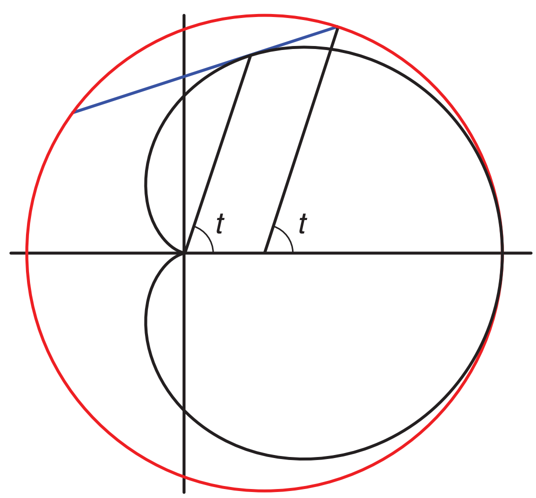


Figure 8. The secant line joining points on the circle with central angles t and $2t$ meets the cardioid at the point with polar angle t.

Back to the Coffee Cup

It turns out that this analysis explains the cardioid in the coffee cup. We can view the caustic as an envelope of lines. As we see in figure 9, if we draw lines emanating from a single point P on the circle and allow them to reflect off the circle (the angle of incidence equalling the angle of reflection), then the cardioid is the envelope of these lines.

If the light source is located at point P, then a beam of

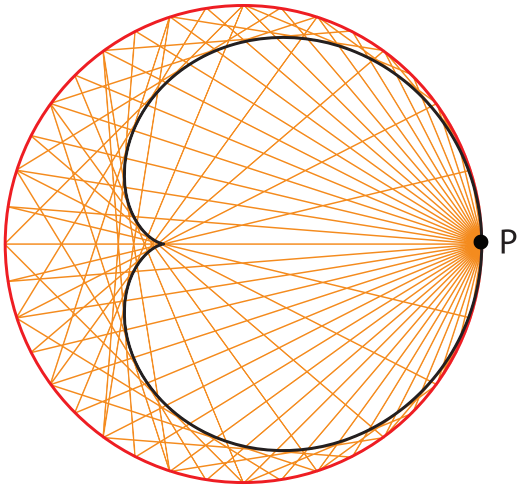


Figure 9. Draw segments from a point on a circle and have them reflect off the second point. The envelope of these lines is a cardioid.

light will reflect off a point Q on the circle and strike the circle again at R (see figure 10). Since arc PQ equals arc QR, arc PR is twice arc PQ. But then segment QR is a line that we would have drawn in the previous construction.

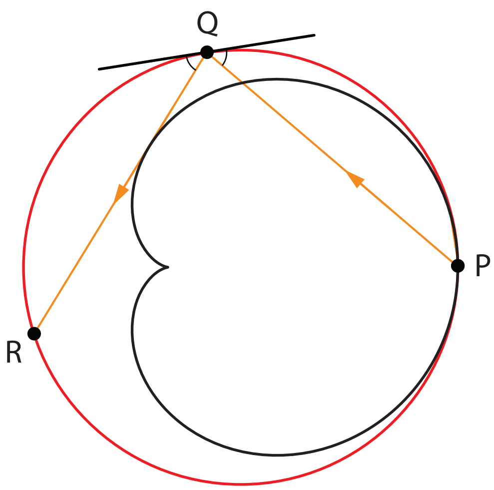


Figure 10. If a ray of light emanating from P strikes the circle at Q and then R, then arc PR is twice arc QR.

The coffee cup example requires one final comment. In reality, the light source will probably not be at the edge of the coffee cup, but rather, it will be far away from the cup. In this case, the rays of light are roughly parallel when they reach the cup. In this case, the curve won't be a cardioid, but its cousin—a *nephroid*. This is the envelope of lines one obtains by joining n and $3n$. In particular, as we see in figure 11, arc QR is twice arc PQ. (So in our numbering, $n = 0$ sits at the point P.)

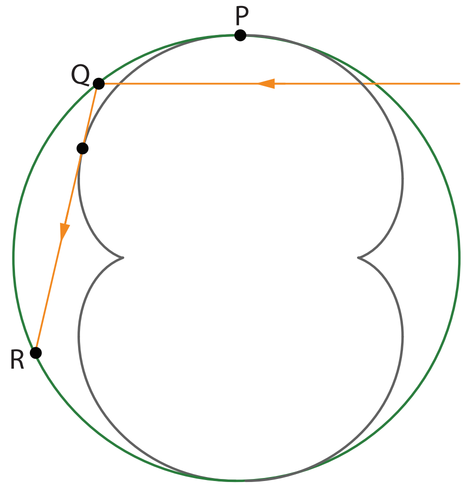


Figure 11. When the light rays come in parallel to one another, the caustic is a nephroid.

Draw Your Own Cardioid

The following page has a circle with 60 numbered points. Connect each number n to the number $2n$ mod 60 to obtain a cardioid. For a little extra fun, try connecting n to $3n$ or $4n$ or $5n$ to see what shapes you obtain.

Flip Book Instructions

The final 12 pages of the article are a printable flip book. Print the pages double-sided. The pages are designed so that the mathematical figure is on one side and the flip book page number is on the reverse side. Cut out each page, put in numerical order, and secure with a binder clip. Flip through the pages and see the animation in action!

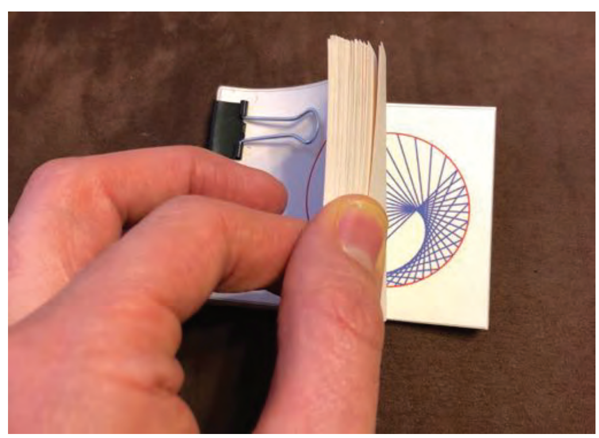
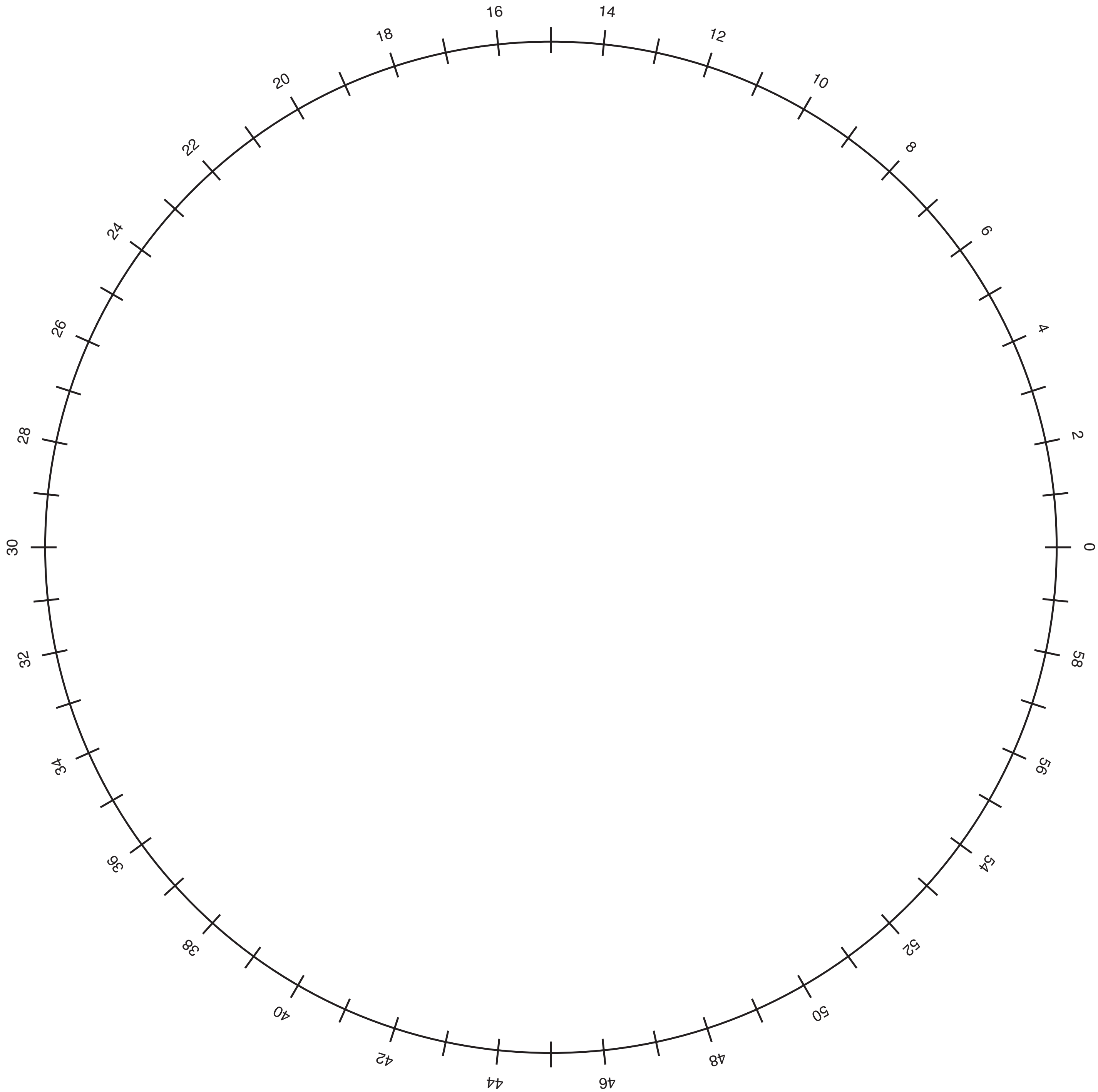
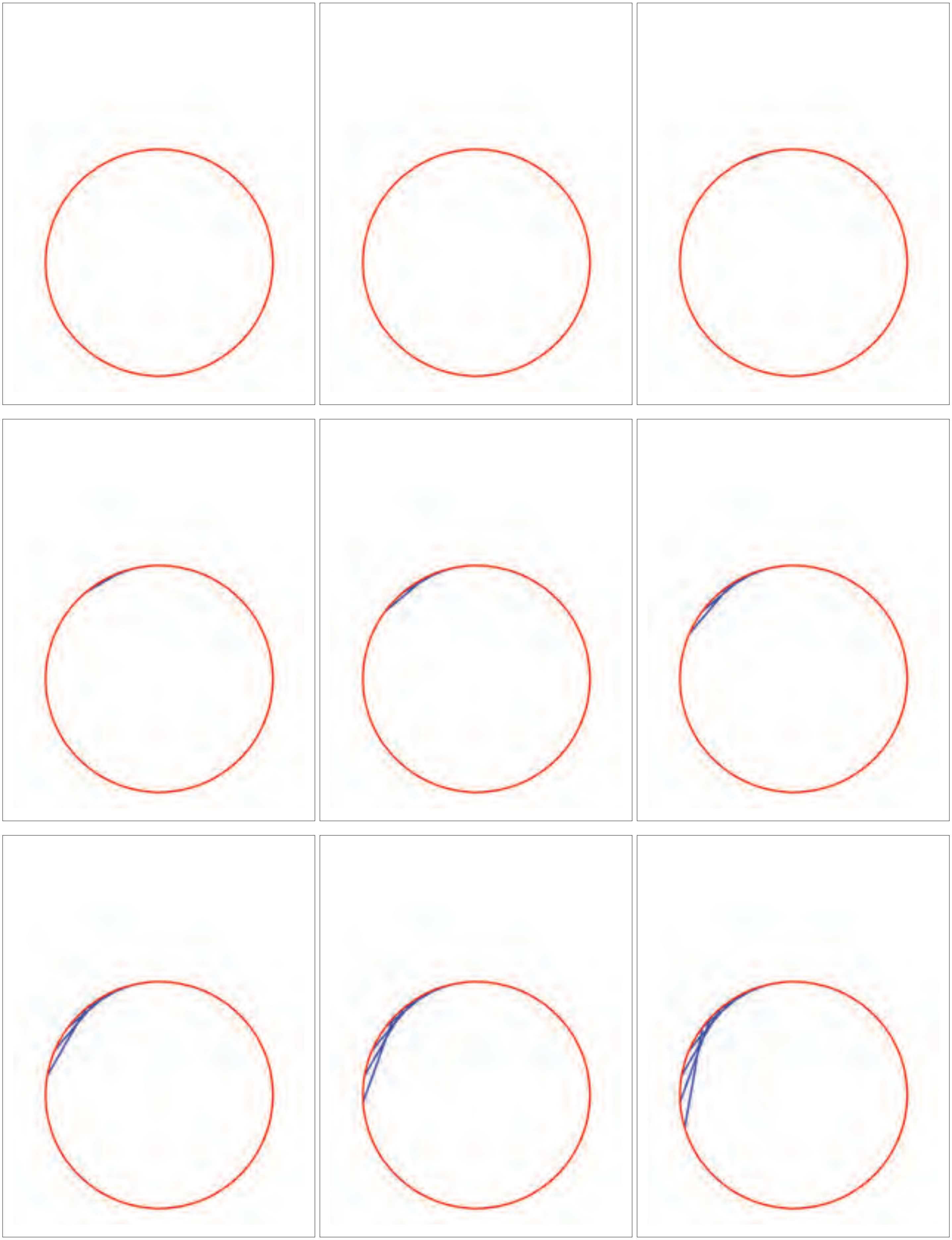
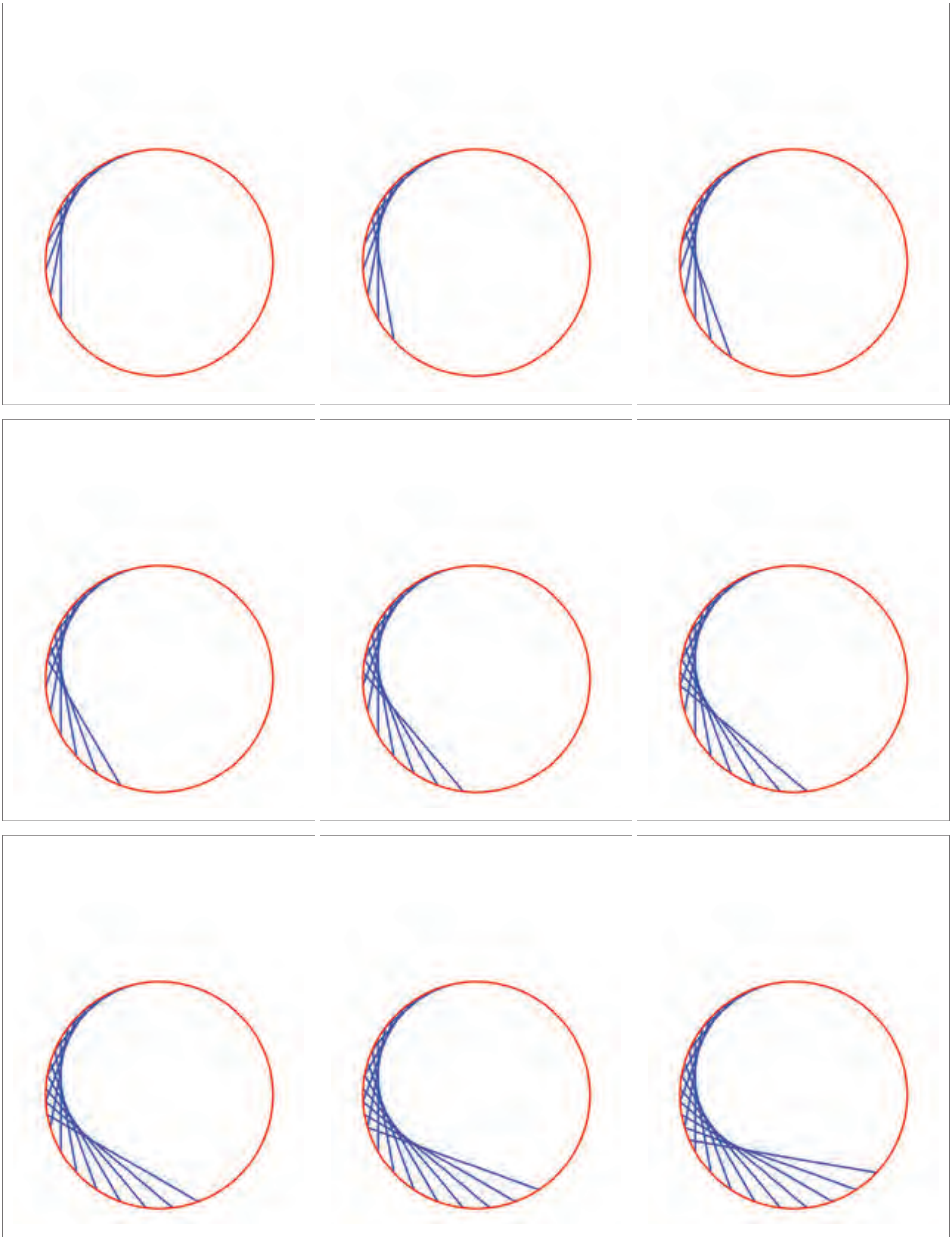


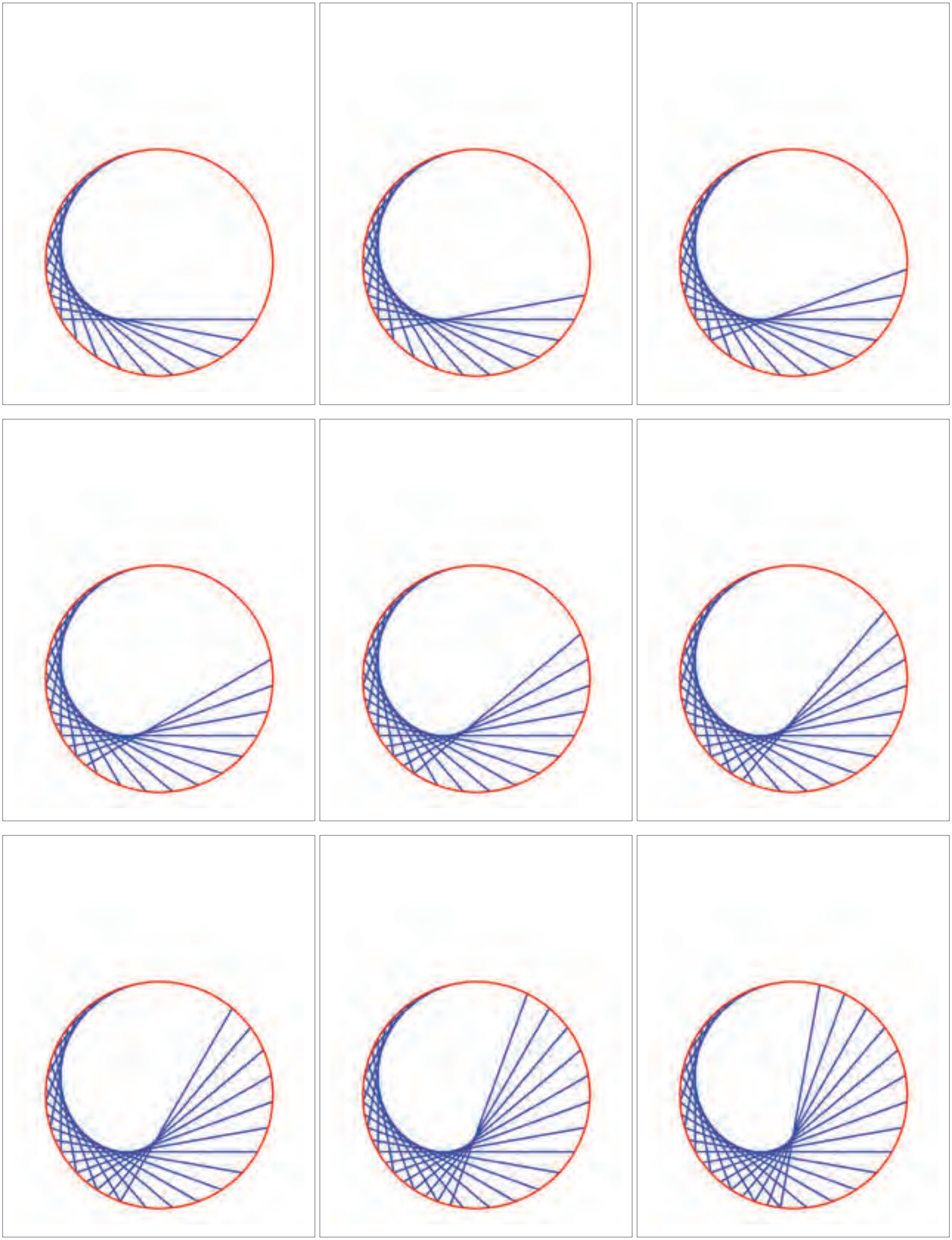
Figure 12. Flip book.

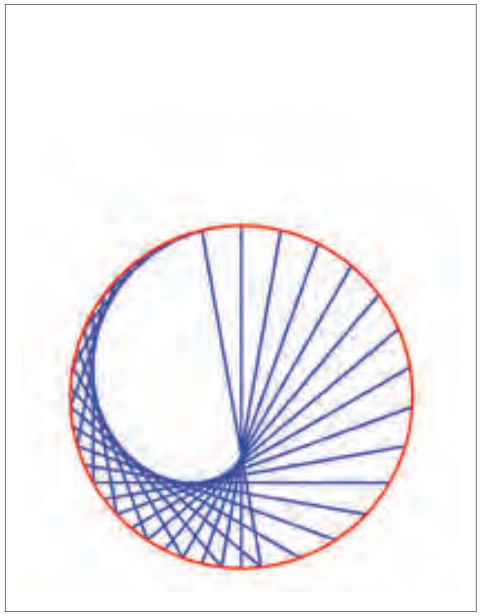
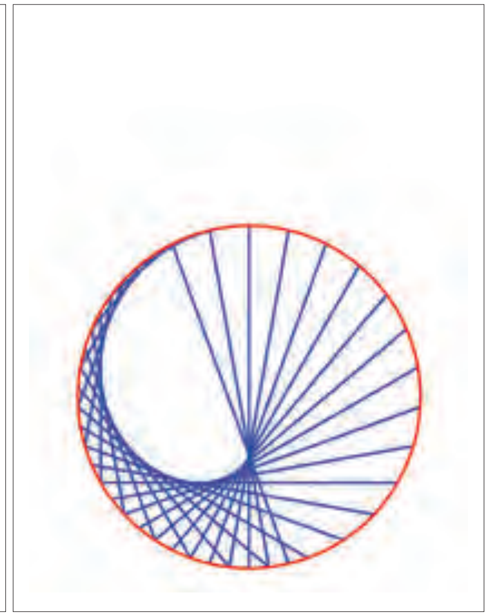
David Richeson is a professor of mathematics at Dickinson College and is the editor of Math Horizons*, the undergraduate magazine of the Mathematical Association of America. You can find him online at* divisbyzero.com *and at* @divbyzero *on Twitter.*

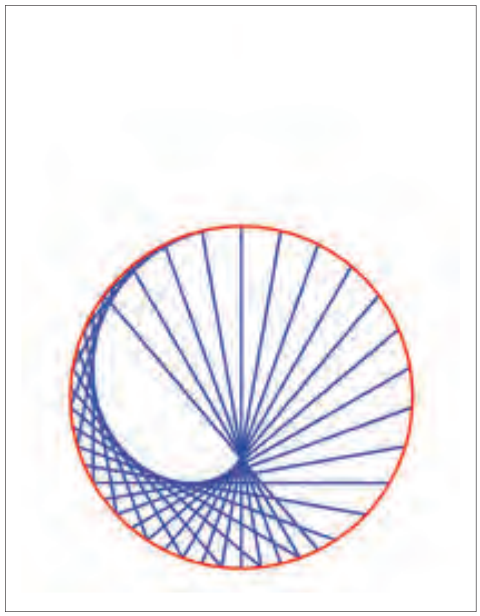

0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58

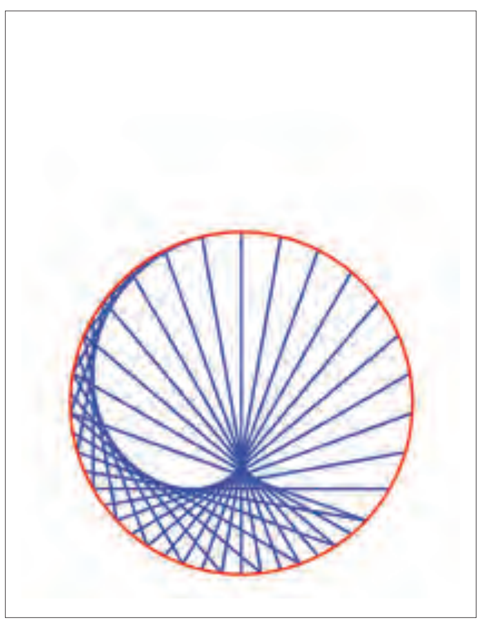
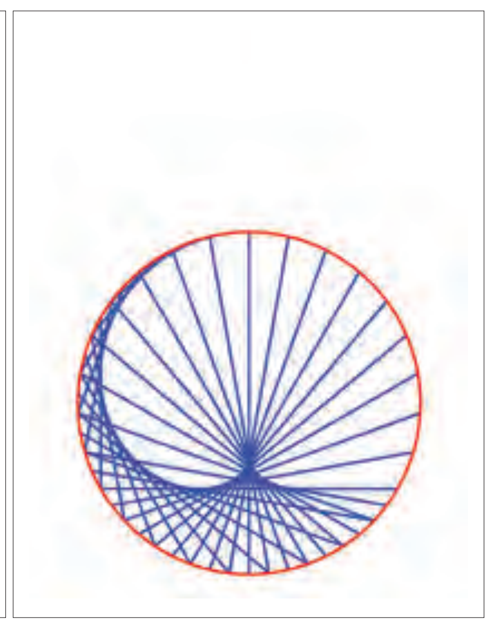


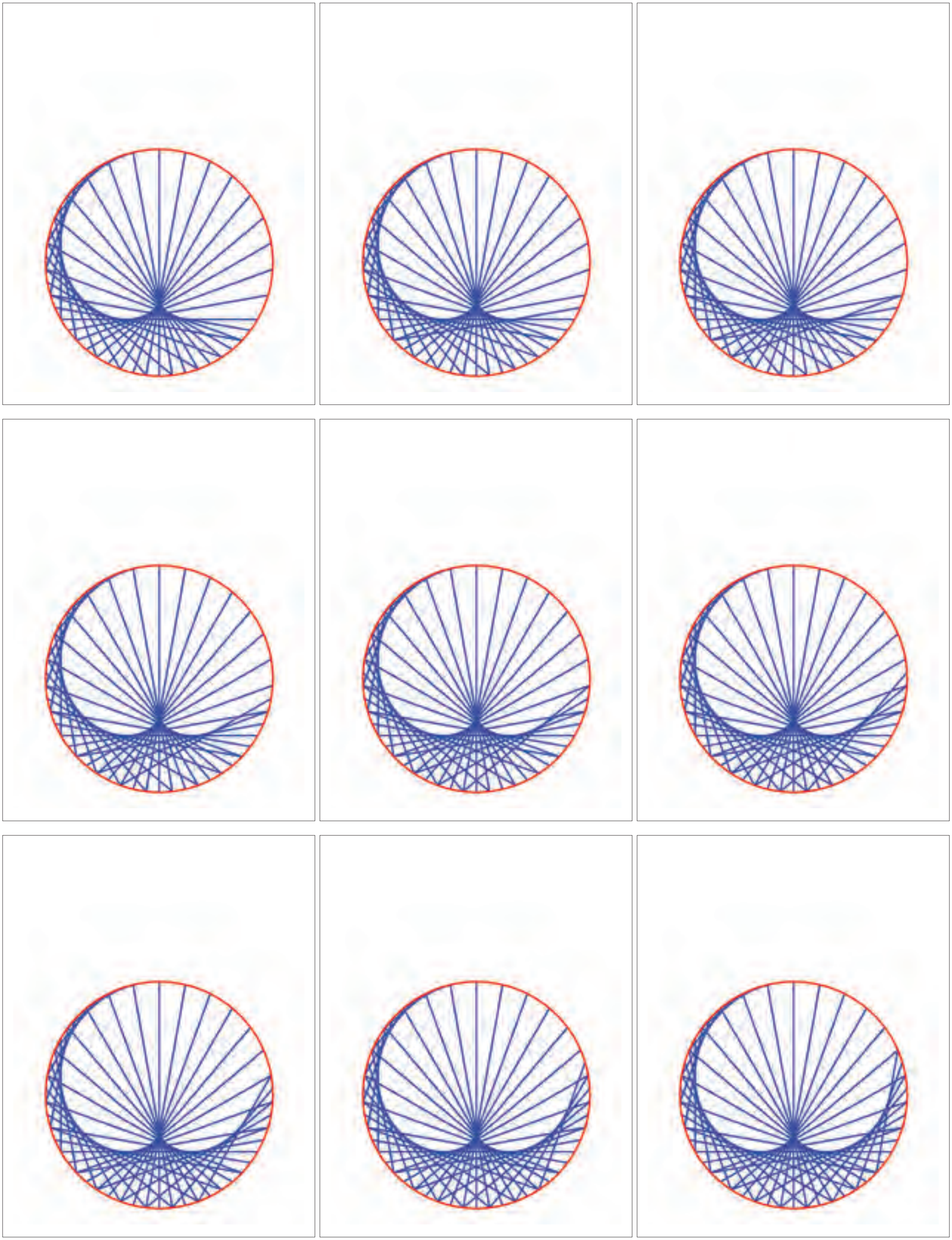


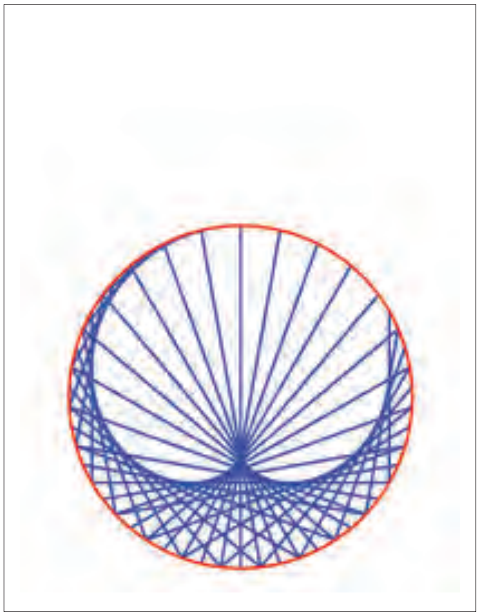
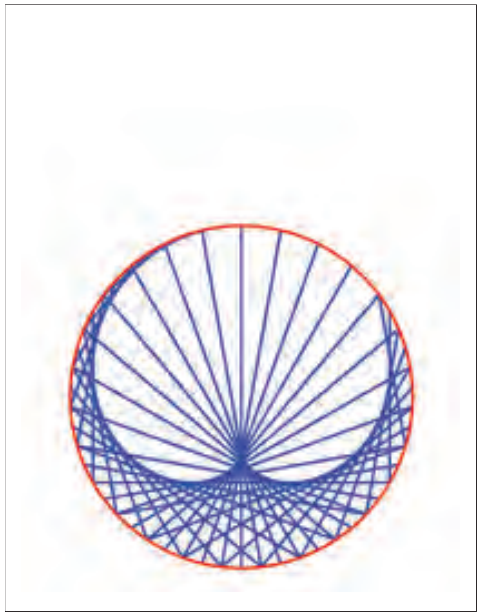


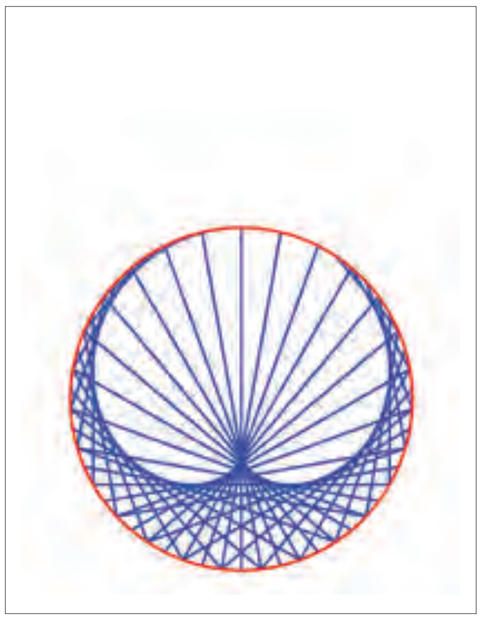








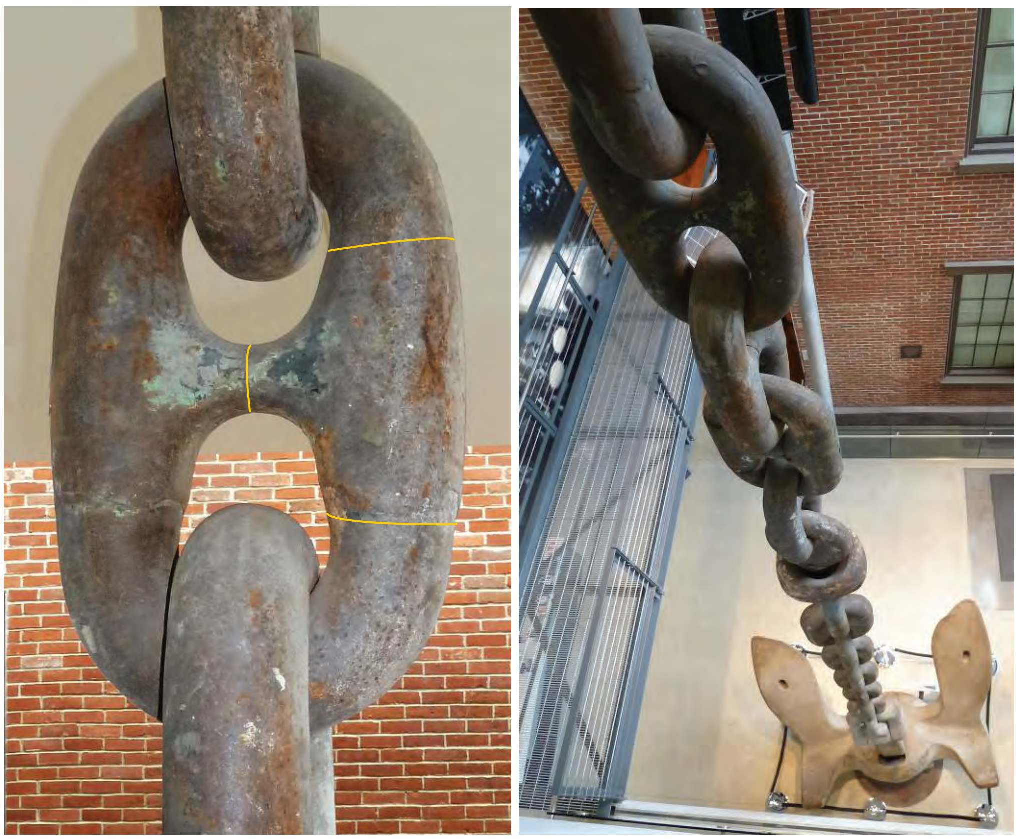


Crocheting Hyperbolic Regular Octagon and Pair of Pants

Daina Taimina

No interesting mathematical topic is self-contained or complete: rather, it is full of "holes", or natural questions and ideas not readily answered by techniques native to the topic. These holes often give rise to connections between the given topic and other topics that seem at first unrelated. – W.P. Thurston[1]

For me, such unrelated topics seemed to be the hyperbolic plane and quantum computing. When in 1997 I first crocheted a hyperbolic plane, the next thing my husband asked me to crochet was a hyperbolic regular octagon with 45-degree interior angles which can be edge-identified (see end of this paper) to form a two-holed torus or *an anchor ring*. Not being a topologist, I did not know about the importance of this figure, but it was interesting to figure out how to crochet it.



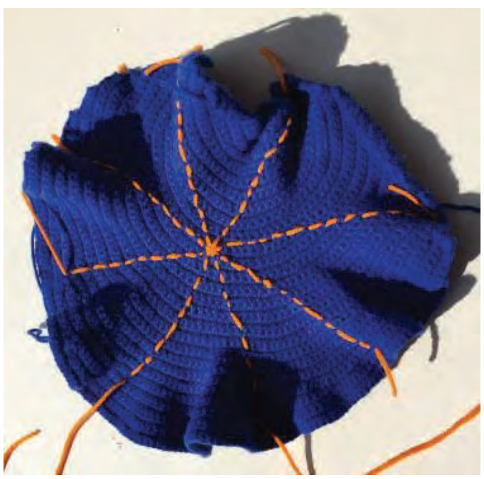
A two-holed torus is sometimes called an "anchor ring" for reasons that are clear in this photo from the Brooklyn Navy Yard.

The 45-degree regular octagon tiles the hyperbolic plane and can also be arranged as a *pair of pants*. Topologically a *pair of pants* is a sphere minus three open disks (one disk removed as a waist and two other disks removed as cuffs for the pants). A pair of pants is indicated by the yellow boundaries in the photo. In hyperbolic geometry, a pair of pants is the smallest building block used for the decomposition of closed surfaces. Recently, I learned that the pair of pants has become of interest in topological quantum field theory[2] and topological quantum computing[3].

However, it was interesting for me to create a pair of pants (and two-holed torus) from a regular hyperbolic octagon with 45-degree interior angles:

Crochet a hyperbolic plane using acrylic yarn. Start with 15 chain stitches and then use single crochet stitches (increase ratio 5 to 6). After eight rows start to shorten rows as shown in a picture. (That is necessary to eliminate unnecessary extra areas on both sides.)

Shortened rows to produce a hyperbolic plane for constructing regular octagon with approximate radius 22 cm.



Central angle divided in eight equal parts.

Constructing octagon. Fold the plane in half and mark the straight line with a thread. Then construct a perpendicular straight line to the first one approximately at the center of the plane. Mark it. Now do two more folds and mark them, so that central angle is divided in eight equal parts.

Make a paper wedge with 45-degree angle and check that central angle is divided equally. Fold the wedge in half to mark its angle bisector.

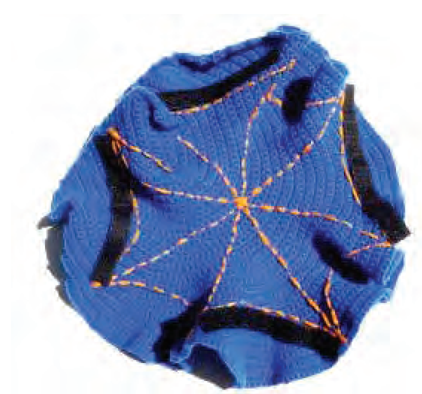
Checking wedge and central angle.

There is really no precise procedure for constructing the 45° octagon needed. After you construct eight 45° angles around the center, mark equal distances in all eight directions (in the model shown in the picture, the distance is approximately 17 cm). Construct one of the sides of the octagon by folding the line between the marks on two adjacent lines and marking it with stitches.

Notice that marking any distance from the center will give you a regular octagon, but not every regular octagon will give the required model: we need one with 45° interior angles. Therefore, after you construct two sides of the octagon, you have to check whether the angle between these two sides is 45°. This can be done more precisely if we instead use our wedge to check that the angle between the side and radius (the line from the center) is 22.5°. Lay one edge of the paper angle along a ray and see how the other edge lines up with the side of the octagon. Working with just two adjacent rays, adjust the distance from the center until the side forms the necessary angles with the rays. Now mark the equal distances from the center along all eight rays accordingly.

Octagon with 45 degree angles

Cut two strips of Velcro closure length of the side of your octagon. In this example the side of the octagon is 20 cm. Attach Velcro at the four sides of the octagon as you can see in the picture.



Fold in and lightly stich excess fabric.

Fasten Velcro strips. Your hyperbolic pants are done!

Hyperbolic pair of pants

More examples can be found in my book:

> *Crocheting Adventures with Hyperbolic Planes: tactile mathematics, art and craft for all to explore*, CRC Press, 2018, 2nd edition.

---

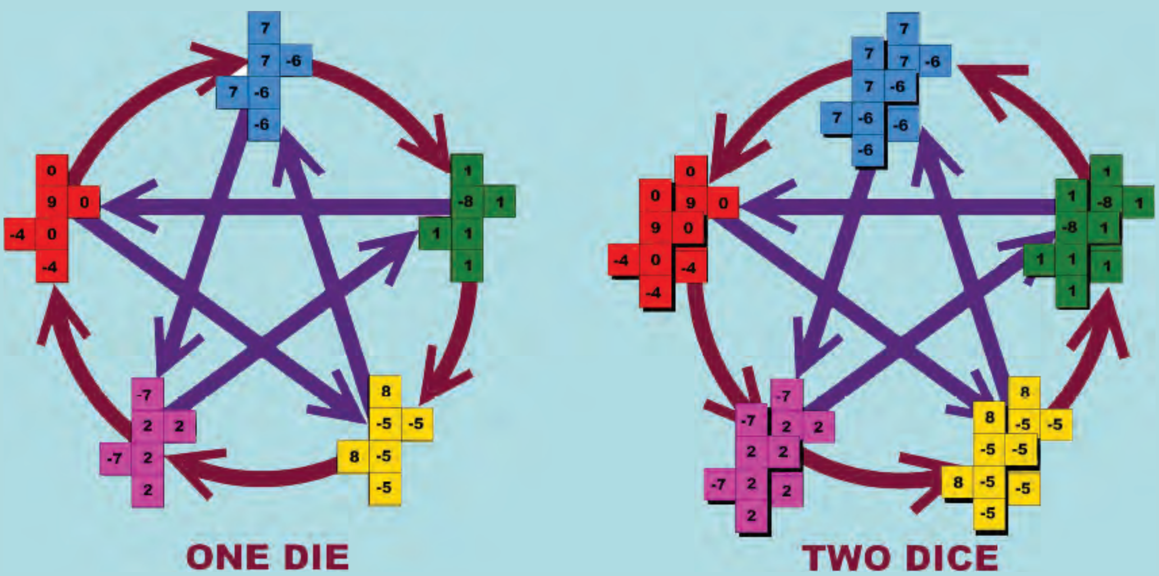
[1] William P. Thurston, *Three-Dimensional Geometry and Topology*, Princeton University Press, 1997, p. ix.

[2] Michael Atiyah, "An Introduction to Topological Quantum Field Theories", *Tr. J. of Mathematics* 21 (1997), p. 1-7.

[2] Slavomir Klimek, Matt McBride, Sumedha Rathnayake, Kaoru Sakai, "The Quantum Pair of pants, Symmetry, Integrability and Geometry: Methods and Applications" *SIGMA 11* (2015), p. 012.

[3] Eric C. Rowell, Zhenghan Wang, "Mathematics of Topological Quantum Computing", 5 Dec 2017, https://arxiv.org/pdf/1705.06206.pdf, p8.

GAMES



Nontransitive Dice | James Grime | Page 104

Don't Cheat!

Spandan Bandyopadhyay

This is a concept for a card game for 3-6 players. It's easy to learn, requires only a few decks of standard cards, and may or may not be centered around cheating. Each player has their own deck and a hand of five cards, and draws and discards a card each round to try to get four-of-a-kind in their hand. And there are a lot of additional rules that definitely do not reward cheating.

I can't tell you that you should try to cheat, but this game would be pretty boring without cheating. Just saying. (Get creative!)

The Rules

Each player brings a standard deck of playing cards to the game. Each deck is shuffled together to make the game deck, which is then dealt out into decks of 52 cards, one deck for each player. Each player then draws five cards from their own deck to form their hand.

The object of the game is to get four-of-a-kind in your hand of five cards. Play occurs in rounds, and a round consists of each player simultaneously drawing a card from their personal deck, choosing a card to discard, and then discarding it facedown in their discard pile. Once everyone has discarded a card, the next round begins. If any deck is exhausted, the player should shuffle their discard pile and use that as their deck. Players can win at any time during a round by placing their hand on the table, face up, so every player can see that they have four-of-a-kind.

There are a few additional rules, as well. If a player believes that another player has three-of-a-kind, they can point to the player and call "Three!" at any time. The player must show their hand to the table, and if they have three-of-a-kind, they must take one of the three, along with the top four cards of their deck, and put them in the calling player's discard pile. Then, they may draw another card from their own pile to replace the one that was given away. If the accused player does not have three-of-a-kind, the accusing player must give the top three cards of his deck away to the accused's discard pile. Note: there is no way to know whether your opponent has three-of-a-kind without cheating, of course, and cheating is not allowed.

Cheating is not allowed, and so there's a penalty for any player cheating. If any player cheats and is caught, they must show the other players their hand and permanently discard a card, choosing a three-of-a-kind card or a two-of-a-kind card, if possible. They are then not allowed to draw a new card to replace the card they discarded, leaving them with a four-card hand. If they are caught cheating again, they lose the game, as you cannot make four-of-a-kind with three cards.

When accusing another player of cheating, special considerations must be taken. The accuser cannot use probability in their accusation – the offending player must be caught in the act of cheating, currently contradicting a rule. An accusation like "The lights went out and when they came back on, you had four aces; you must be cheating," is not valid because there exists a small probability that they acquired four aces legitimately. The accuser is allowed to call out actions that are violating unwritten, common-sense rules, however. An accusation like "You're not allowed to put your hand cards into your pocket and

bring them out again" is valid. Writing out these unwritten rules would limit the creativity of cheating techniques, which are still not allowed, by the way, so they will remain unwritten. However, the existence of these unwritten rules may require an Arbiter – see below.

Variations

There are plenty of variations of this game that are worth trying once you get the hang of the game. Here's a few that I've thought up, but there's many ways you can play this that I haven't thought of yet.

Assigning an Arbiter

In any game with purposefully general rules about cheating and accusations, baseless accusations can get very overwhelming, and baseless denials of legitimate accusations may be thrown out even faster. When these begin to interfere with gameplay, the assignment of an arbiter is useful. The arbiter sits out of the game and is the final judge of what is reasonable to assign as cheating and what is not, and when a player is making too many baseless accusations. A new arbiter each game makes each game different and balances the scales a bit. Of course, a player could bribe an arbiter, but that's cheating.

Additional players

Don't Cheat is meant to be played by 3-6 people, as the decks get unwieldy after that point, but a 2-player version could be possible, as well as a game adapted for more players. Two-player games are much more difficult because each player can focus on tactics employed by the other; they don't have to split their attention. As for the 7+ player game, grouping into teams should work well. If there are 9 players, for example, split into teams of three. No team will win without all three of its members having four-of-a-kind.

Friends!

Everybody has a friend. If there's an even number of players, your friend is the person across from you. If there's an odd number of players, your friend is the player directly to the right of you. The additional rule is that accusing your friend of cheating is against the rules. This leads to interesting cooperative opportunities – but if you decide to call your friend on cheating, they lose a card for cheating, and you lose a card for cheating by breaking that rule.

Playing with fewer cards

It's not difficult for someone to win the main game without any cheating at all – getting four of a kind in a deck of fifty-two cards is very possible. However, if you want to make the game harder, you can limit the deck sizes to thirty-nine or even twenty-six cards. This would give each deck a moderately unlikely chance to contain any set of four at all and would increase the difficulty for advanced players. It should definitely not be used to encourage more cheating.

Acknowledgements

Brainstorming and feedback help by Brenda Castro, Neil Glikin, Ravin Sajnani, and my wonderful family.

Scrabble® Seven-letter Words

Tom Bessoir and Joshua Pines
February 2018

"It's a damn poor mind that can only think of one way to spell a word."
- Andrew Jackson

Abstract

Scrabble® may be the world's most popular word game and is played in more than 30 different languages. Over 150 million Scrabble sets have been sold in over 120 countries. This paper analyzes various statistics for the English-language edition of the game of Scrabble. Specifically we investigate the mathematics of the seven-tile starting racks and seven-letter words, and determine the likelihood that a starting rack can make a seven-letter word.

Introduction

Scrabble[1] is a multiplayer word game where players compete for the highest score by using letter tiles to make words crossword-style. It was developed by Alfred Mosher Butts[2], an American architect, in 1938 as a variant of his game Lexiko. Lexiko is played without a board, similar to dominoes. When Butts created Scrabble, he added a board.

The board contains 225 squares arranged in a 15 by 15 grid. There is a pool of 100 tiles. Each tile fits on a square and contains a letter and a point value. Each player has a rack that holds seven tiles randomly chosen from the pool. A play consists of placing tiles from your rack onto the board to form new words. The point values of each tile in the new word or words created are added to that player's score. If all seven tiles are used in one turn, this is known as making a "bingo" and the player receives a 50 point bonus. Making bingos is a key strategy of advanced players.

Certain squares on the board have bonus properties when newly covered by a tile. The four types of bonus squares are shown in Table 1.

Table 1 - Bonus Square Counts

bonus square	count
Double Letter Score	24
Triple Letter Score	12
Double Word Score	17
Triple Word Score	8

Bonus squares are only mentioned for completeness; this paper will not be addressing tile placement.

Alfred Butts determined a point value for each letter based on its frequency in written English words. He eventually settled on the tile distribution shown in Table 2, which is still in use today.

Table 2 – Tile Distribution

letter	point value	tile count
A	1	9
B	3	2
C	3	2
D	2	4
E	1	12
F	4	2
G	2	3
H	4	2
I	1	9
J	8	1
K	5	1
L	1	4
M	3	2

letter	point value	tile count
N	1	6
O	1	8
P	3	2
Q	10	1
R	1	6
S	1	4
T	1	6
U	1	4
V	4	2
W	4	2
X	8	1
Y	4	2
Z	10	1
?	0	2

There are two blank tiles in the pool which are the equivalent of "wild cards" and which may be used to represent any letter. In this paper, a blank tile is represented by a question mark.

In an interview Stefan Fatsis said, "... the game is all about math. There are 100 tiles, 98 letters and two blanks. It's all about combinations, and they are mathematical."[3] In the present paper, we investigate the mathematics of seven-letter letter words and starting racks. We are not going to be concerned with tile point values and word scores.

The Dictionary

Acceptable words are determined as those contained in some chosen reference source. Casual players may simply agree to use any dictionary at hand. Many players agree to use the Official Scrabble Players Dictionary or OSPD. Tournament and club players in the United States and Canada use the more comprehensive Official Tournament and Club Word List. In this paper, we will use the third and current edition of this reference source: OTCWL2014, also known as OWL2014.

We refer to OWL2014 as the "dictionary," even though it is just a word list and does not contain pronunciations, etymologies, or definitions. OWL2014 contains 24,029 seven-letter words.

Alphabetical Bingos

The dictionary was checked for bingos with their letters in alphabetical order. Two lists of the 24,029 bingos in the dictionary were created. In the second list, the letters in each bingo were sorted alphabetically left to right. A line by line comparison was performed of the original list of bingos against the list of alphabetized bingos, checking for matches.

The only bingos that have their letters in alphabetical order are:

```
BEEFILY
BILLOWY
```

Palindromic Bingos

The dictionary was checked for bingos that are palindromes. A palindrome is a word or phrase that reads the same forwards and backwards. An example of a palindromic phrase is “Never odd or even.” Examples of some common words that are palindromes include `CIVIC`, `NOON`, and `SEXES`.

Two lists of the 24,029 bingos in the dictionary were created. In the second list the letters in each bingo were reversed. A line by line comparison was performed of the original list of bingos against the list of reversed bingos, checking for matches.

The only palindromic bingos are:

```
DEIFIED
HALALAH
REIFIER
REPAPER
REVIVER
ROTATOR
SEMEMES
```

The commonly known seven-letter palindrome, `RACECAR`, is considered two separate words in our dictionary (`RACE` and `CAR`), so it is not in the above list.

Bingos with Duplicated Letters

The 24,029 bingos in the dictionary were checked for words with duplicated letters. Each letter in a word may appear once, twice, three times, or four times. There are no bingos with more than four instances of the same letter.

For this analysis, our algorithm used four counters to track the number of single, double, triple, and quadruple letter occurrences. For each bingo, the letters were sorted alphabetically left to right. This sorting brings the repeated letters together, making it easier to count how many single, double, triple, and quadruple letters occur. The eleven possible letter-repetition patterns are shown in Table 3.

Table 3 – Letter-Repetition Patterns

letter count				examples	
single	double	triple	quadruple	bingo	alphabetized bingo
7				CINEMAS	ACEIMNS
5	1			BLUFFED	BDEFFLU
3	2			TUITION	IINOTTU
1	3			GRAMMAR	AAGMMRR
4		1		LOWBALL	ABLLLOW
2	1	1		EELLIKE	EEEIKLL
	2	1		ALFALFA	AAAFFLL
1		2		LESSEES	EEELSSS
3			1	ASSISTS	AISSSST
1	1		1	REFEREE	EEEEFRR
		1	1		

There are no bingos with both a triple and a quadruple letter.

The letter repetitions in each alphabetized bingo were examined and the appropriate counters were incremented. The number of bingos for each of the eleven patterns is presented in Table 4.

Table 4 – Bingo Counts by Letter-Repetition Patterns

letter count				number of bingos
single	double	triple	quadruple	
7				7,940
5	1			11,040
3	2			3,413
1	3			195
4		1		989
2	1	1		399
	2	1		15
1		2		5
3			1	24
1	1		1	9
		1	1	0
			total	24,029

The number of bingos containing various repetitions of each letter was examined. For example, how many bingos contained exactly three A's?

We created 104 counters - four counters for each of the 26 letters. These counters tracked the number of single, double, triple, and quadruple occurrences of that letter. For each of the 24,029 bingos, the letters were again sorted alphabetically left to right, bringing the repeated letters together and making it simpler to count how many singles, pairs, triplets, and quadruplets occur for each letter. The letter repetitions in each alphabetized bingo were examined and the appropriate counters were incremented.

For example, the word `ALFALFA` would be alphabetized as `AAAFFLL` and the "triple `A`," "double `F`," and "double `L`" counters would each get incremented. The results are presented in Table 5.

Table 5 – Bingo Counts by Letter Repetitions

letter	number of bingos			
	single	double	triple	quadruple
A	9,328	1,727	184	1
B	3,063	347	27	
C	5,029	476	6	
D	5,453	660	74	2
E	11,888	3,426	431	11
F	1,877	315	12	
G	4,181	497	65	2
H	3,689	146	3	
I	9,828	1,545	30	
J	428	9		
K	2,100	101		
L	7,155	1,017	41	
M	3,947	356	25	
N	7,773	1,020	63	
O	6,907	1,318	71	10
P	3,965	476	28	
Q	308			
R	9,337	1,465	36	
S	10,253	2,278	238	5
T	7,141	1,083	67	
U	5,534	366	6	1
V	1,385	45		
W	1,827	48	2	
X	528	3		
Y	2,592	66		
Z	538	99	4	1

Bingos That Require Blanks

Some of the 24,029 bingos in the dictionary require one or more blanks because of the distribution of letter tiles in the pool. For example `KNOCKED` requires a blank to be used for one of the two `K`'s, `MAXIMUM` and `MINIMUM` both require a blank to be used for one of the three `M`'s, and `ZIZZLES` requires two blanks to be used for two of the three `Z`'s.

To calculate the words that require blank tiles, we started once again by sorting the letters alphabetically left to right for each bingo, bringing the repeated letters together. The number of times each letter appears in the word was checked against the tile pool to see if the pool contains enough tiles of that letter. If the pool doesn't contain enough tiles of that letter, then one or more blank tiles would be needed to make that bingo. Each alphabetized bingo was examined to determine how many blanks were required to make that word with the distribution of letter tiles in the pool.

Table 6 - Bingos That Require Blanks

blanks required	number of bingos
0	23,707
1	317
2	4
3	1
total	24,029

The last line in Table 6 indicates that there is a seven-letter word in the dictionary that requires more blanks than are available in the tile pool - and there is: the word is `PIZZAZZ`!

Letter Combinations That Can Not Bingo

The question of letter combinations that do not occur on any of the 24,029 bingos in the dictionary was explored. Were there any two-letter combinations that could be drawn from the tile pool which would prevent making a bingo? Were there any three-letter combinations?

A list of all 351 two-letter combinations: `AA`, `AB`, `AC`, etc. was created. Letter combinations that can not be drawn from the tile pool were eliminated from the list. For example, `JJ` was eliminated, as there is only one `J` in the tile pool. After eliminating 5 combinations (`JJ`, `KK`, `QQ`, `XX`, and `ZZ`), there remained 346 two-letter combinations that can be drawn. For each of these two-letter combinations, the list of bingos was checked until a word that contains those two letters was found. A matching word was found for every pair of letters that can be drawn, so there are no two-letter combinations that will prevent making a bingo.

Next the three-letter combinations were considered. Similarly, a list of all 3,276 three-letter combinations was created and those that can not be drawn from the tile pool were eliminated. After eliminating 139 combinations, there remained 3,137 three-letter combinations that can be drawn. For each of these three-letter combinations, the list of bingos was checked until a word that contains those three letters was found. If no matching word was found, those three letters block the possibility of making a bingo. The 201 three-letter combinations that prevent making a bingo are shown in Table 7.[4]

Table 7 – Three-Letter Bingo Blockers

A J Q	A J X	A Q X	B B V	B C Q	B D Q	B F J	B F Q	B G Q	B J P
B J Q	B J V	B K Q	B M Q	B P Q	B Q V	B Q W	B Q X	B V W	B V X
B V Z	B X Z	C F Q	C G Q	C J Q	C J V	C J X	C Q V	C Q W	C Q X
C W Z	C X Z	D G Q	D J Q	D J V	D J X	D J Z	D K X	D Q V	D Q W
D Q X	D Q Y	E J Q	F G Q	F H Q	F H V	F J K	F J Q	F J V	F J X
F J Z	F K Q	F K V	F M Q	F N Q	F O Q	F P Q	F P V	F P Z	F Q T
F Q V	F Q W	F Q X	F Q Z	F V X	F V Z	F W Z	F X Z	G G X	G J Q
G J V	G K X	G L Q	G M Q	G Q V	G Q W	G Q X	G Q Z	G V Z	G X Z
H H Q	H H X	H J L	H J Q	H J V	H J X	H J Y	H J Z	H K Q	H K X
H P Q	H Q T	H Q V	H Q W	H Q X	H Q Z	H V V	H V X	H W X	H X Z
I Q W	J K Q	J K V	J L X	J M Q	J M V	J M W	J M X	J P Q	J P V
J P W	J P X	J P Z	J Q R	J Q S	J Q T	J Q V	J Q W	J Q X	J Q Y
J Q Z	J R X	J S X	J T Z	J U W	J V V	J V W	J V X	J V Z	J W W
J W X	J W Z	J X Y	J X Z	J Y Y	K M X	K P Q	K Q V	K Q X	K Q Z
K T X	K U V	K V W	K V X	K X Y	K X Z	L Q V	L Q W	L Q X	L X Z
M M Q	M P Q	M Q V	M Q W	M Q X	M V V	M V W	M W Z	M X Z	N Q V
N Q W	N Q Z	O Q W	O Q Z	P Q V	P Q W	P Q X	P Q Y	P Q Z	P V X
P V Z	P W Z	P X Z	Q R X	Q S X	Q V V	Q V W	Q V X	Q V Z	Q W W
Q W X	Q W Z	Q X Y	Q X Z	Q Y Y	Q Y Z	R X Z	S X Z	T X Z	U W W
U X Z	V V W	V V X	V V Z	V W W	V W X	V W Z	V X Z	W W Z	W X Z
X Y Z									

Starting Racks

A starting rack consists of seven randomly drawn tiles from the pool of 100 tiles. We calculated the total number of possible starting racks.

We used combinations instead of permutations since the order in which the tiles were chosen doesn't matter. The formula for choosing k items from a set of n items without replacement is:

$$C(n, k) = n*(n-1)*...*(n-k+1) / [k*(k-1)*...*1] \qquad (1)$$

By using factorials, this formula can be written as:

$$C(n, k) = n! / [(n-k)! * k!] \qquad (2)$$

For a starting rack, this is simply the number of ways to choose seven tiles from the 100-tile pool.

$$n(\text{starting racks}) = C(100,7) = 100! / (93! * 7!) = 16,007,560,800 \qquad (3)$$

Blanks and Starting Racks

The 100-tile pool contains two blank tiles.

Number of starting racks with no blanks:

$$n(\text{starting racks}_{\text{blanks}=0}) = C(98,7) = 13{,}834{,}413{,}152 \quad (4)$$

Number of starting racks with 1 blank:

$$n(\text{starting racks}_{\text{blanks}=1}) = C(2,1)*C(98,6) = 2{,}105{,}236{,}784 \quad (5)$$

Number of starting racks with both blanks:

$$n(\text{starting racks}_{\text{blanks}=2}) = C(98,5) = 67{,}910{,}864 \quad (6)$$

Vowels and Consonants and Starting Racks

As shown in Table 2, there are 42 vowel tiles, 56 consonant tiles, and two blank tiles in the pool.

Let n(starting racks $_{bvc}$) be the number of starting racks with b blanks, v vowels, and c consonants. The formula for calculating them is:

$$n(\text{starting racks}_{bvc}) = C(2, b)*C(42, v)*C(56, c) \quad (7)$$

The number of starting racks for all combinations of blanks, vowels, and consonants are tabulated in Tables 8 through 10.

Table 8 - Possible Starting Racks with 0 Blanks

vowels	consonants	combinations
7	0	26,978,328
6	1	293,764,016
5	2	1,310,028,720
4	3	3,102,699,600
3	4	4,216,489,200
2	5	3,288,861,576
1	6	1,363,674,312
0	7	231,917,400
	total	13,834,413,152

Table 9 - Possible Starting Racks with 1 Blank

vowels	consonants	combinations
6	0	10,491,572
5	1	95,274,816
4	2	344,744,400
3	3	636,451,200
2	4	632,473,380
1	5	320,864,544
0	6	64,936,872
	total	2,105,236,784

Table 10 - Possible Starting Racks with 2 Blanks

vowels	consonants	combinations
5	0	850,668
4	1	6,268,080
3	2	17,679,200
2	3	23,866,920
1	4	15,426,180
0	5	3,819,816
	total	67,910,864

The totals for each table provide a check against the previous rack calculations for various numbers of blanks and ensure that all of the cases were covered.

Unique Starting Racks

There may be several ways of selecting seven tiles from the pool that result in the same starting rack. As an example, there are multiple ways of drawing the `AAAAAAA` starting rack from the 100-tile pool, since there are nine `A` tiles.

We defined a rack to be "unique" if it has a unique collection of letters and blanks, independent of their order. To simplify determining if a rack is unique, the seven tiles were kept in alphabetic order with the blanks on the right. There may be multiple ways of drawing each unique rack from the 100-tile pool, since there are multiple tiles for many letters.

In the above example of drawing `AAAAAAA`, the number of ways of drawing this starting rack is simply the number of ways of choosing seven tiles from the pool of the nine `A` tiles:

$$n(\text{starting rack}_{\text{AAAAAAA}}) = C(9,7) = 36 \qquad (8)$$

We wanted to determine the number of unique starting racks that can be created by drawing tiles from the tile pool. All of our above results on starting racks were obtained using combinatorics,

but to determine the number of unique starting racks, we wrote a C program with a recursive subroutine.

An alphabetized list of the tiles in the 100-tile pool was created. The blanks, represented by question marks, are at the end of the list:

```
A A A A A A A A A B B C C . . . V V W W X Y Y Z ? ?
```

The program has seven pointers. The first pointer represents the first tile in the rack, the second pointer represents the second tile in the rack, etc. The first pointer starts by pointing to the first tile in the list, second pointer points to the second tile in the list, etc. So initially, the seven pointers produce:

```
AAAAAAA
```

This is the first unique rack.

The seventh pointer was advanced through the list of tiles in the pool. Each time the pointer advances to a tile that has a different letter from the previous tile, a new unique rack was created, and the count of unique racks was incremented. When the seventh pointer reaches the 100th tile in the list, the sixth pointer was advanced through the list until it encounters a tile that has a different letter from the previous tile it was pointing at. The seventh pointer was then repositioned to the tile to the right of where the sixth pointer is now pointing, creating a new unique rack, and the count of unique racks was incremented.

The process repeated with the seventh pointer advancing through the list counting the number of unique racks. Every time the seventh pointer reaches the last tile in the list, the sixth pointer was advanced through the list to a tile that has a different letter and the process repeats.

When the sixth pointer reaches the 99th tile in the list (the seventh pointer is at the 100th tile), the fifth pointer was now advanced through the list to a tile that has a different letter, the sixth and seventh pointers were repositioned to the two tiles to the right of the new fifth-pointer tile and the count of unique racks was incremented. The process begins again with the seventh pointer advancing through the list.

When a pointer reaches the other pointers at the end of the list, the next lower pointer advances to a tile that has a different letter. Eventually the first pointer will advance to the 94th tile, the process will stop and each unique rack will have been counted.

Our program counted 3,199,724 unique starting racks.

Starting Racks and Bingos

After analyzing the properties of the dictionary, tile distribution, and starting rack composition, we then analyzed the relationship between the seven tiles on the starting racks and the seven-letter words in the dictionary.

Starting Racks That Can Make the Most Bingos

We determined which starting racks made the most bingos. A list of the 24,029 bingos in the dictionary was created and the letters in each bingo were again sorted alphabetically left to right.

In addition to letter tiles, we have the two blank tiles, which can be used for any letter. There are additional racks that can be made by substituting blanks for one or two letters for each alphabetized bingo. There are three ways to add one or two blanks. First, each unique letter in the bingo can replaced with a blank. Second, each letter that occurs two or more times in the bingo can be replaced with double blanks. Third, each unique pair of different letters in the bingo can be replaced with two single blanks. We present an example to clarify this concept.

The bingo `EELLIKE` when alphabetized yields `EEEIKLL`. The following racks can be made by substituting blanks for one or two tiles:

no blanks	one single blank	double blanks	two single blanks
EEEIKLL	?EEIKLL	??EIKLL	?EE?KLL
	EEE?KLL	EEEIK??	?EEI?LL
	EEEI?LL		?EEIK?L
	EEEIK?L		EEE??LL
			EEE?K?L
			EEEI??L

We added blanks as described above to create the expanded rack list and again sorted each rack alphabetically left to right, with the blanks on the far right.

Some of these racks can not be drawn because they have more occurrences of a letter than there are tiles of that letter in the pool. This happens because there are words in the dictionary with letters occurring more times than there are tiles for that letter, and blanks must be used for those letters. When our algorithm did not replace these letters with blanks, it generated a rack that could not be drawn from the tiles in the pool. For example the word `PUZZLES` can be made from a hypothetical rack of `ELSUZZ?`, but this rack can not be drawn because there is only a single `Z` tile in the pool. We eliminated those hypothetical racks that can not be drawn from the tile pool.

Then the list was sorted alphabetically top to bottom. Now any starting rack that can be used to make more than one bingo will be repeated on adjacent lines in the list (e.g. `FLATCAR` and `FRACTAL` will both be sorted to produce `AACFLRT`). As the number of times each line is duplicated corresponds to the number of words in the dictionary created from that starting rack, we counted the repeated lines to see which starting racks match the most words.

The starting racks that can make the most bingos based on the number of blanks are shown in Table 11.

Table 11 – Starting Racks That Can Make the Most Bingos

blank tiles	starting racks	bingos
0	AEINRST	9
1	AEIRST?	70
2	AERST??	406

Starting Racks That Can Bingo

We wanted to know the probability of drawing a starting rack that can make a bingo. The brute force solution would be to check each of the over 16 billion possible starting racks to see if those tiles could be used to make any of the 24,029 bingos in the dictionary. We devised a more efficient approach.

We were not concerned with the words themselves; we were only concerned with the combinations of letters that make up the word. First, we started with the list of bingos and used it to build a list of all the unique starting racks that can make a bingo. Then we determined how many ways each of these unique starting racks could be drawn from the tile pool.

A list of the 24,029 bingos in the dictionary was created and the letters in each bingo were again sorted alphabetically left to right. Then the list was sorted alphabetically top to bottom so that any combination of letters that can be used to make more than one bingo will be repeated on adjacent lines in the list. By eliminating these duplicate lines from the list, we determined that there are 20,134 unique letter combinations needed to make all 24,029 bingos.

We now had to account for the two blank tiles. To each unique letter combination, we added zero, one, or two blanks as described above to create an expanded rack list. We applied the same technique used previously of sorting the tiles left to right with the blanks on the far right. Next the list was sorted alphabetically top to bottom and the duplicates were eliminated to leave only the unique racks.

Now that we had a list of all unique blanks-added racks that can produce a bingo, the next step was to count the number of possible ways each of these unique racks can be drawn from the tile pool. For example if our rack is `EEEIKLL`, we compute the number of ways to draw three `E`'s from the 12 available `E`'s, draw one `I` from the nine available `I`'s, draw one `K` from the one available `K`, and draw two `L`'s from the four available `L`'s. We multiply these numbers together to obtain the possible ways to draw the rack.

$$n(\text{starting rack}_{\text{EEEIKLL}}) = C(12,3) * C(9,1) * C(1,1) * C(4,2) = 11{,}880 \qquad (9)$$

We created a list of these racks and the number of possible ways they can be drawn from the tile pool. For any racks that could not be drawn due to the tile distribution, this number was zero. These impossible racks were eliminated and the list of unique bingo racks was what remained.

$$n(\text{unique bingo racks}) = 120{,}828 \qquad (10)$$

We summed all the counts for the possible ways to draw each of these unique bingo racks to determine:

$$n(\text{starting racks that can make a bingo}) = 2{,}068{,}621{,}350 \qquad (11)$$

Dividing the number of starting racks that can make a bingo by the total number of starting racks yields:

$$2{,}068{,}621{,}350 / 16{,}007{,}560{,}800 = 0.1292 \text{ (to four decimal places)} \qquad (12)$$

Thus 12.92% of randomly drawn starting racks can make a bingo.

Blanks and Starting Racks That Can Bingo

Our list of starting racks that can make a bingo were analyzed with respect to the number of blank tiles on the rack.

Table 12 – Blank Analysis

blank tiles	combinations	Percentages of	
	starting racks that can bingo	all starting racks that can bingo	all starting racks
0	1,075,220,956	52.0%	6.72%
1	938,048,008	45.3%	5.86%
2	55,352,386	2.7%	0.35%
totals	2,068,621,350	100.0%	12.92%

The totals provide a check with our previous work.

Monte Carlo Simulation of Starting Racks That Can Bingo

Since the authors were quite surprised by the high probability that a starting rack could make a bingo, we decided to check our results against a computer simulation. We created a program which randomly selected seven tiles from the 100-tile pool one million times and checked to see if those seven tiles could make a bingo. This Monte Carlo simulation was run ten times with different random seeds. Our result was 1,292,224 bingos for ten million starting racks.

This confirms our calculated 12.92% of starting racks can make a bingo.

Conclusions

There are 24,029 bingos in the dictionary. There are 20,134 unique letter combinations that will make all 24,029 bingos. Two of the bingos have their letters in alphabetical order and seven of the bingos are palindromes.

The most common letter-repetition patterns for bingos are five unique letters plus a pair, followed by seven unique letters. The least common letter-repetition pattern for bingos is two sets of triple letters (only five words). The dictionary contain 33 seven-letter words with a quadruple letter.

Examination of Table 5 reveals that the dictionary contains no bingos with two or more `Q`'s and no bingos with three or more `J`'s, `K`'s, `V`'s, `X`'s, or `Y`'s. The only letters that occur four times in a bingo are `A`, `D`, `E`, `G`, `O`, `S`, `U`, and `Z`.

Most bingos can be made without using a blank tile, but 317 bingos require one blank tile and four bingos require both blank tiles. One of the bingos in the dictionary, `PIZZAZZ`, can not be made with the tiles available, as it would require three blanks.

There are 16,007,560,800 ways to choose a seven-tile starting rack from the 100-tile pool. A starting rack will contain no blanks 86.4% of the time, one blank 13.2% of the time, and both

blanks 0.4% of the time. A starting rack will contain all consonants 1.5% of the time, and all vowels 0.2% of the time. The three most common starting racks are three vowels and four consonants (26.3%), two vowels and five consonants (20.6%), and four vowels and three consonants (19.4%).

The starting racks that can make the most bingos are `AERST??` (406 bingos) and `EIRST??` (340 bingos). The starting rack that can make the most bingos without a blank is `AEINRST` (9 bingos). Of all the racks that can bingo, the most likely to be drawn is `AEEINRT`, which has 1,154,736 ways it can be drawn. This rack can make three bingos, namely `ARENITE`, `RETINAE`, and `TRAINEE`. Of all the racks that can bingo, the least likely to be drawn are `BBCCK??` and `CCHHK??`, each with only one way to draw these tiles. These racks can make the bingos `BIBCOCK` and `CHACHKA` respectively.

The answer to the question which started us on all of this Scrabble research surprised us. The probability that a randomly drawn starting rack can make a bingo is an unexpectedly high 12.92%.

A future paper will present our research using tile point values to analyze the scoring potential of starting racks and bingos.

Acknowledgments

The author order is alphabetical reflecting equal contribution by both authors and in deference to the alphabetical sorting used in many of the algorithms. We thank Donna Atwood, Rod Bogart, Robert Follek, David Klein, and Robyn Stukalin for comments on early drafts of this paper.

Notes

1. Scrabble® is a registered trademark of Hasbro, Inc. in the United States and Canada. Outside of the United States and Canada, it is a trademark of J.W. Spear & Sons Limited, a subsidiary of Mattel, Inc.
2. The history of Scrabble and Lexiko is from “Word Freak: Heartbreak, Triumph, Genius, and Obsession in the World of Competitive SCRABBLE Players” by Stefan Fatsis (2001)
3. The Pennsylvania Gazette, September/October 2001; http://www.upenn.edu/gazette/0901/fatsissidebar.html
4. Although not a true bingo blocker, `VYZ` can only make one bingo. All other three-letter combinations not shown in Table 7 can make at least four bingos. The only bingo containing `VYZ` is `ZYZZYVA`, which is a South American weevil and also happens to be the last word in many English language dictionaries.

Nontransitive Dice for Three Players

James Grime

1 Three Unusual Dice

Here is a game you can play with a friend. It is a game for two players, with a set of three dice. These dice are not typical dice however, because instead of having the values 1 to 6, they display various unusual values.

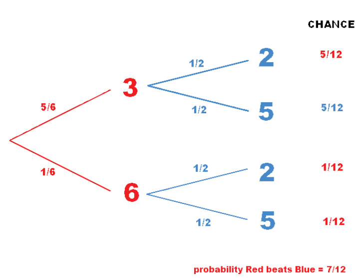
The game is simple: Each player picks a die. The two dice are then rolled together and whoever gets the highest value wins.

The game seems fair enough. Yet, in a game of, say, ten rolls, you will always be able to pick a die with a better chance of winning - no matter which die your friend chooses. And you can make these dice at home right now.

Here is the set of three special dice:

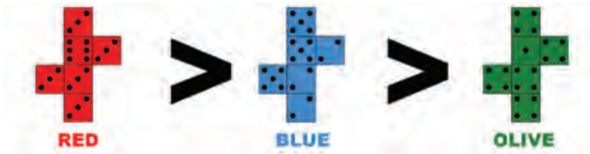
We say A beats B if the probability of die A beating die B is greater than 50%.

It's simple to show that the Red die beats the Blue die by way of a tree diagram:



From the diagram we see Red beats Blue with a probability of $\frac{7}{12}$. This is greater than 50% so Red is the better choice here.

Similarly, it can be shown that Blue beats Olive with a probability of $\frac{7}{12}$. So we can set up a winning chain where Red beats Blue, and Blue beats Olive.



Using this information it would be perfectly reasonable to expect, therefore, that Red beats Olive. If this is true we call the dice 'transitive'.

However, this is not the case. In fact, bizarrely, Olive beats Red with a probability of $\frac{25}{36}$. This means the winning chain is a circle - like a game of 'Rock, Paper, Scissors'.

This is what makes the game so tricky because, as long as you let you opponent pick first, you will always be able to pick a die with a better chance of winning.

2 Double Whammy

After a few defeats your friend may have become suspicious, but all is not lost. Once you've explained how the dice beat each other in a circle, challenge your friend to one more game.

This time you will choose first, in which case your opponent should be able to pick a die with a better chance of winning. But let's increase the stakes, and increase the number of dice. This time each player rolls two of his chosen die, so that the player with the highest total wins.

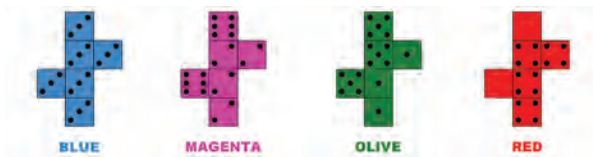
Maybe using two dice means your opponent has just doubled their chances of winning. But not so because, amazingly, with two dice the order of the chain flips!

In other words, the chain reverses so the circle of victory now becomes a circle of defeat - allowing you to win the game again!

3 Efron Dice

The paradoxical nature of nontransitive dice goes back to 1959 and to the Polish mathematicians Hugo Steinhaus and Stanislaw Trybula [4].

However, the remarkable reversing property is not true for all sets of nontransitive dice. For example, here is a set of four nontransitive dice introduced by Martin Gardner in his column Mathematical Games in 1970 [2]. This set is known as 'Efron Dice' and was invented by the American statistician Brad Efron:



Here, the dice form a circle where Blue beats Magenta, Magenta beats Olive, Olive beats Red, and Red beats Blue, and they each do so with a probability of $\frac{2}{3}$.

Usiskin and Trybula independently showed [7], [6] that it was always possible to set up a nontransitive system of m n-sided dice, and showed that the weakest winning probability has a bound. It is not possible for all winning probabilities to exceed this bound, but it is possible for all winning probabilities to be greater than, or equal to, this bound.

For six-sided dice, the set of three dice above achieve this bound. Using a different number of sides the greatest bound for three dice is the Golden Ratio $\varphi = 0.618\ldots$. This theoretical bound increases as the number of dice increases, and converges to $\frac{3}{4}$.

Efron Dice achieve the bound for four dice of $\frac{2}{3}$. Unfortunately, they do not possess remarkable reversing property when you double the number of dice - while some of the probabilities reverse, others do not.

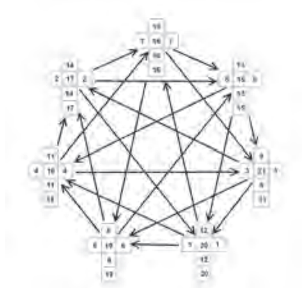
It is said the billionaire American investor Warren Buffett is a fan of nontransitive dice. When he challenged his friend Bill Gates to a game, with a set of Efron dice, Bill became suspicious and insisted Warren choose

first. Maybe if Warren had chosen a set with a reversing property he could have beaten Gates - he would just need to announce if they were playing a one die or two dice version of the game after they had both chosen.

4 Three Player Games

I wanted to know if it was possible to extend the idea of nontransitive dice to make a three player game - a set of dice where two of your friends may pick a die each, yet you can pick a die that has a better chance of beating both opponents - at the same time!

It turns out there is a way. The Dutch puzzle inventor M. Oskar van Deventer came up with a set of seven nontransitive dice, with values from 1 to 21. Here two opponents may each choose a die from the set of seven, and yet there will be a third die with a better chance of beating each of them. The probabilities are remarkably symmetric with each arrow on the diagram illustrating a probability of $\frac{5}{9}$.



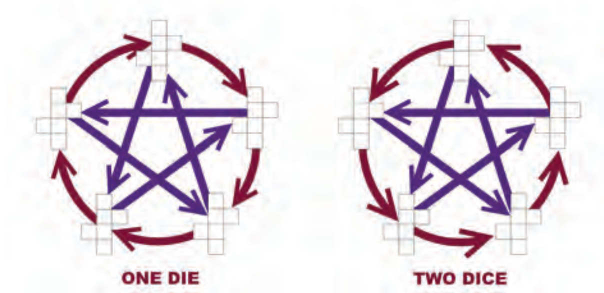
This means we can play two games simultaneously, however beating both players at the same time is still a challenge. The probability of doing so stands at around 39%.

This set of seven dice form a complete directed graph. In the same way, a four player game would require 19 dice. A direct construction of this set was not known until 2016 when Angel and Davis devised a direct construction for any tournament of any number of dice [1].

However, I began to wonder if it was possible to exploit the revering property of some nontransitive dice to design a slightly different three player game, one that uses fewer than seven dice.

5 Grime Dice

My idea for a three player game required a set of five dice that contained two nontransitive chains. When the dice were doubled one chain would remain in the same order, while the second chain would reverse. This way, choosing a one or two dice version of the game will allow you to play two opponents at the same time, no matter which dice they pick.



After a small amount of trial and error, I devised the following set of five nontransitive dice:

These dice appeared to be the best set of five I could find. I have written about them before and they became known as Grime Dice.

For one die games we get the following chains:

All winning probabilities here are equal to, or greater than, $\frac{5}{9}$ with an average winning probability of 63% - I will leave the calculations to the interested reader. Notice the first chain is ordered alphabetically, while the second chain is ordered by word-length.

You can also find nontransitive subsets of dice. For example, the Red, Blue and Olive dice are a copy of the original set of three nontransitive dice that I describe above, complete with the same winning probabilities and reversing property.

For two dice games we get the following chains:

An unfortunate consequence of Red, Blue and Olive having the reversing property is that, when we double the dice, the first chain (the outside circle) reverses order, while the second chain (the inside pentagram) stays the same - with one exception.

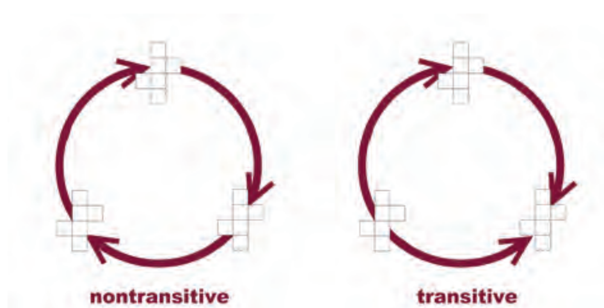
However, the probability of this exception is very close to 50% (it's $\frac{625}{1296}$). Meanwhile, the average of all other winning probabilities is 62% - much higher than Oskar Dice - and so, in practice, the three player game still works.

It's quite nice that this set of five contained three dice with their own reversing property. However, I admit, the exception continued to niggle at me. I wanted to know if there was a set of five nontransitive dice, with the desired properties, with no exceptions - or was this set really as close as we could get.

6 Finding a new set of Grime Dice

I enlisted the help of a computer, and the invaluable help of my friend Brian Pollock, to search for sets of five nontransitive dice. The computational challenge of working out all sets of five dice, and their chains, was a large one - so we devised a test.

Three dice can either form a diagram with all three arrows in the same direction, which we call a nontransitive chain, or with only two arrows in the same direction, which we call a transitive chain.



We wanted to create a set of five nontransitive dice, with two nontransitive chains, such that, when doubled, one chain stays the same, and the other chain reverses order.

This will mean that, for any subset of three dice, if they form a nontransitive chain singly, they will form a transitive chain when doubled. Alternatively, if they form a transitive chain singly, they will form a

nontransitive chain when doubled. If a chain remains transitive or nontransitive when the dice are doubled, then we say the set has failed the test.

There are 10 subsets of three dice from a set of five. Each subset needs to pass the test. Furthermore, if all subsets pass the test, we have found a valid set of five dice with the desired properties.

The application of this test allowed us to reject sets without the desired property with less calculation.

In the first instance, we only considered dice using the values 0 to 9. Sets of dice that allow draws would be rather unsatisfactory, so after excluding draws, no set of five dice passed the test.

Only a few sets of four dice passed the test, which simply turned out to be the original Grime Dice with one of the dice missing. This proved that Grime Dice really are the best set of five dice using the values 0 to 9, without draws.

7 Dice with higher values

Naturally, the next thing to try were dice with higher values. Keeping the criteria of no draws, the first success found used the values 0 to 13.

A : 4 4 4 4 4 9
B : 2 2 2 7 7 12
C : 0 5 5 5 5 10
D : 3 3 3 3 8 13
E : 1 1 6 6 6 11

There were two such sets using the values 0 to 13, with the second set being only a slight variation of the above. These were also the only sets of five with the desired properties that uses consecutive numbers.

I was delighted with this success, but the winning probabilities for this set were weaker than Grime Dice. The average winning probability is about 59%, lower than Grime Dice. So we continued our search, to find a set with stronger winning probabilities.

The winning probabilities slowly increased as we increased the values on the dice. Here is one of the strongest sets of five dice using the values 0 to 17:

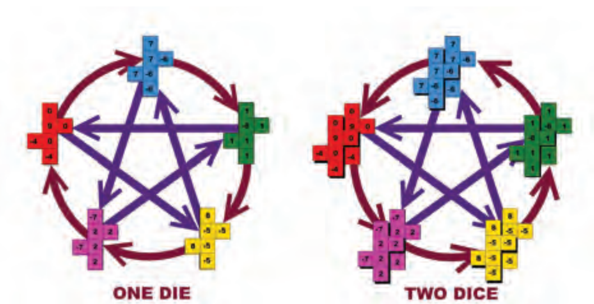
A : 4 4 8 8 8 17
B : 2 2 2 15 15 15
C : 0 9 9 9 9 9
D : 3 3 3 3 16 16
E : 1 1 10 10 10 10

Increasing the dice values after this point did nothing to improve the winning probabilities. Since the

numbers are no longer consecutive there is enough space for the values to change without changing the winning probabilities, meaning this set can appear repeatedly in slightly different forms. The investigation for better sets had plateaued.

For aesthetic reasons, I decided to subtract 8 from all sides of the above dice, making a set of New Grime Dice using the values from -8 to 9:

Like the original Grime Dice, this set makes two nontransitive chains, one with the colours listed alphabetically, the other with the colours listed by word length. When doubled, the alphabetical chain remains in the same order, while the chain ordered by word length flips.



In single dice games, New Grime Dice have the exact same winning probabilities as Original Grime Dice. When the dice are doubled, New Grime are generally slightly weaker than Original Grime Dice. The average winning probability for New Grime Dice is 60.4%, that's 0.7% lower than Original Grime Dice. Crucially however, all winning probabilities are now over 50%, allowing for a true three player game as follows:

Invite two opponents to pick a die each, but do not volunteer whether you are playing a one die or two dice version of the game. No matter which dice your opponents pick, you will always be able to pick a die to beat each opponent. If your opponents pick two dice that are consecutive alphabetically, play the one die version of the game. If your opponents pick two dice that are consecutive by word-length, use the two dice version of the game.

8 A Gambling Game

But, can we expect to beat the two other players at the same time? Well, we have certainly improved the odds, with the average probability of beating both opponents now standing around 44% - a 5% improvement over Oskar Dice. So, if the odds of beating two players isn't over 50% then how do we win? Consider the following gambling game:

Challenge two friends to a dice game, where you will play your two opponents at the same time. If you lose you will give your opponent 1. If you win your opponent gives you $1. So, if you beat both players at the same time you win $2; if you lose to both players you lose $2; and if you beat one player but not the other then your net loss is zero. You and your friends decide to play a game of 100 rolls.

If the dice were fair then each player will expect to win zero - each player wins half the time and loses half the time.

However, with Oskar Dice, you should expect to beat both players 39% of the time, and lose to both players 28% of the time, which will give you a net profit of $22.

But even better, with New Grime Dice, you should expect to beat both players 44.1% of the time, but only lose to both players 23.6% of the time, giving you an average net profit closer to $41! (And possibly the loss of two former friends).

I invite you to try out these games yourself, and enjoy your successes and failures!

References

[1] L. Angel, M. Davis, A Direct Construction of Non-Transitive Dice Sets. *Journal of Combinatorial Designs.* (2016) 10.1002/jcd.21563.

[2] M. Gardner, Mathematical games: The paradox of the nontransitive dice and the elusive principle of indifference, *Sci. Amer.* **223** (Dec. 1970) 110114.

[3] R. P. Savage Jr., The paradox of nontransitive dice, *Amer. Math. Monthly* **101** (1994) 429436, http://dx.doi.org/10.2307/2974903.

[4] H. Steinhaus, S. Trybula, On a paradox in applied probabilities, *Bull. Acad. Polon. Sci.* **7** (1959) 6769.

[5] S. Trybula, On the paradox of three random variables, *Zastos. Mat.* **5** (1960/1961) 331332.

[6] S. Trybula, On the paradox of n random variables, *Zastos. Mat.* **8** (1965), 143-154.

[7] Z. Usiskin, Max-Min Probabilities in the Voting Paradox, *Annal. Mat. Stat.* **35** (June 1964), 857-862.

Deck Building Games with Playing Cards

Fred Henle ⟨fredhenle@gmail.com⟩
http://fredhenle.net

for G4G13

Abstract

Here are two new deck building games that may be played with an ordinary 52-card deck of playing cards (appropriately for **G4G13** as such a deck has **4** suits of **13** cards).

1 Introduction

I have long enjoyed playing deck building games such as Dominion or Ascension. In a deck building game, each player begins with a small individual deck of low level cards and over the course of the game acquires more (and better) cards. Some of the enjoyment comes from beautiful artwork and themes, but the game mechanics themselves are fun, and so I wanted to invent a similar game that could be played using an ordinary deck of cards, and would therefore be much more portable. In this article I will describe two such games, with variations.

2 Structure and Common Assumptions

Both games assign a numeric value to each card in a typical way, ace through king representing integer values between 1 and 13.

These games are best played at a table; each player should have space for an individual deck (face down), a discard pile, and separate pile (the "vault") of cards removed from the game by the player. In the center of the table will be the remaining cards (the "store") in three stacks with only the top card of each stack face up.

In both games each player's starting deck contains ten cards. The player draws up to a hand of five cards from his or her deck (shuffling the discard pile when necessary to replenish the deck), then plays cards in an attempt to "buy" a new card from the store by creating a numeric expression whose value is exactly that of the desired card. Cards that have been bought are immediately replaced by turning over the exposed card below. Under certain circumstances cards may also be removed from the player's deck and "banked" in the vault. The games end when the store has an empty stack.

	hearts	clubs	diamonds	spades	
1	A♥	A♣	A♦	A♠	ace
2	2♥	2♣	2♦	2♠	two
3	3♥	3♣	3♦	3♠	three
4	4♥	4♣	4♦	4♠	four
5	5♥	5♣	5♦	5♠	five
6	6♥	6♣	6♦	6♠	six
7	7♥	7♣	7♦	7♠	seven
8	8♥	8♣	8♦	8♠	eight
9	9♥	9♣	9♦	9♠	nine
10	10♥	10♣	10♦	10♠	ten
11	J♥	J♣	J♦	J♠	jack
12	Q♥	Q♣	Q♦	Q♠	queen
13	K♥	K♣	K♦	K♠	king

Figure 1: Card values.

Where the games differ is in the composition of the initial decks, the expression logic, and the scoring rules.

3 RPN

This game is named for *Reverse Polish Notation*. Suits represent arithmetic operations, and so each card has both an operand (value) and an operator, and the 52 cards in the deck are therefore functionally unique.

3.1 Setup

Use one deck for two players, two decks for three or four players, three decks for five or six players, et cetera. For your first game, if you have enough spare decks from which you can borrow cards, each player should start with the same set of ten cards: A♥ 2♥ 3♥ A♣ 2♣ A♦ 2♦ 3♦ A♠ 2♠. Otherwise, just randomly deal each player six hearts and four clubs. See Sec. 5.6 for more ideas on starting decks.

Shuffle each player deck of ten cards. Shuffle the main deck and split into three stacks with only the top card of each stack face up. Determine who should go first.

value	cards	RPN	expression
1	7♠ J♣ 9♦ K♥	7 11- 3× 13+	$((7-11)\times 3)+13$
2	K♥ J♣	13 11-	$13-11$
4	K♥ 9♦ J♣ 7♠	13 3× 11- 7÷	$((13\times 3)-11)\div 7$
5	9♦ J♣ K♥	3 11- 13+	$(3-11)+13$
	9♦ K♥ J♣	3 13+ 11-	$(3+13)-11$
6	K♥ J♣ 9♦	13 11- 3×	$(13-11)\times 3$
9	7♠ J♣ K♥	7 11- 13+	$(7-11)+13$
	7♠ K♥ J♣	7 13+ 11-	$(7+13)-11$
10	7♠ 9♦ J♣	7 3× 11-	$(7\times 3)-11$
11	9♦ J♣ K♥	9 11- 13+	$(9-11)+13$
	9♦ K♥ J♣	9 13+ 11-	$(9+13)-11$

Figure 2: Sample expressions in RPN

3.2 The Turn

When it is your turn, if you haven't already done so, draw five cards from your deck into your hand. You may then buy a card from the store by forming an expression, using at least two of the cards in your hand, whose value is that of the card. The first card in the expression is simply its value, regardless of suit. Each subsequent card in the expression combines its value with an operation: hearts for addition, clubs for subtraction, diamonds for multiplication, and spades for division. Note that you must always use at least two cards in your expression.

In this game, cards in your hand with composite values may be treated as any of their factors greater than one (only aces are allowed to have the value 1). Therefore the most versatile cards are queens, since they may be treated as any of $\{2, 3, 4, 6, 12\}$.[1] See Fig. 2 for the possible values and derivations for the partial hand K♥ J♣ 9♦ 7♠.

As soon as you buy a card from the store, the card below it in the stack is turned over. You then have an opportunity to buy it or one of the other face up cards with the cards remaining in your hand.

At the end of your turn, discard any unused cards.

3.3 Additional Rules

At any point during your turn, you may discard a heart from your hand (i.e., a heart that hasn't already been played that turn) to draw a replacement card from your deck. At any point during your turn, you may discard a club from your hand to bank another yet-unplayed card: adding it to your vault, where it shall remain until the game ends.

3.4 Scoring

The game ends when the store has at least one empty stack. Add up your victory points from the cards in your hand, discard pile, deck, and vault. Hearts are not worth any victory points; clubs are worth 1 victory point each, diamonds are worth 2, and spades are worth 3. See Sec. 5.7 for a more advanced scoring variant.

[1] This only applies to cards in your hand, not the store; you may only buy a queen with an expression whose value is 12.

4 Golomb

I named this game after Solomon Golomb, mathematician and long time member of the G4G community.

4.1 Setup

Use one deck of cards for two players, but add a second deck for three or more. For your first game, try a starting deck of A♥ 2♥ 4♥ 8♥ J♥ in one suit (hearts, clubs, diamonds, or spades) and 3♣ 6♣ 10♣ Q♣ K♣ in a different suit. For these starting decks, only use as many suits as there are players (but keep unused suits in the main deck). For future games, each player should take an entire suit and choose five cards to pass to the player to the right, then choose three out of the thirteen cards (five from the player to the left and eight original cards) to return to the main deck.

Shuffle each player deck of ten cards. Shuffle the main deck and split into three stacks with only the top card of each stack face up. Determine who should go first.

value	cards	expression
1	K♣ Q♣	(13 − 12)
2	Q♣ 10♣	(12 − 10)
	Q♣ 10♣ J♥	(12 + 10) ÷ (11)
3	J♥ 8♥	(11 − 8)
	K♣ 10♣	(13 − 10)
	J♥ 8♥ K♣ Q♣	(11 − 8) × (13 − 12)
4	Q♣ J♥ 8♥	(12) ÷ (11 − 8)
6	Q♣ 10♣ J♥ 8♥	(12 − 10) × (11 − 8)
8	8♥ K♣ Q♣	(8) × (13 − 12)
9	K♣ 10♣ J♥ 8♥	(13 − 10) × (11 − 8)
11	J♥ K♣ Q♣	(11) × (13 − 12)

Figure 3: Sample expressions in Golomb

4.2 The Turn

When it is your turn, draw from your deck into your hand until you have five cards in your hand (you may already have cards in your hand from the previous turn). Group your cards by suit. If you have two cards of the same suit, you may use either as its value, or you may use both together as their sum or difference. If you have more than two you can extend this in a natural way. Another way to put this is that you can, using the cards of a single suit, express any number than may be obtained by adding and/or subtracting the values of those cards. So, if you have A♥ 2♥ 4♥ in the same suit, you may express the values 1, 2, and 4 with individual cards, or $3 = 4 - 1$, $5 = 4 + 1$, $6 = 4 + 2$, or $7 = 4 + 2 + 1$. This logic is reminiscent of the *Golomb ruler*, hence the name of the game.

Given values derived from different suits, you may use multiplication and/or division to combine those values. See Fig. 3 for the possible values and derivations for the hand 8♥ J♥ 10♣ Q♣ K♣. Note that you must always use at least two cards, so there's no way to get the value 12 even though that hand includes a queen.

4.3 Additional Rules

As soon as you buy a card from the store, the card below it in the stack is turned over. You then have an opportunity to buy it or one of the other face up cards with the cards remaining in your hand. If you were unable to play any cards this turn, you may discard as many as you like from your hand, but your turn ends there. If you played between two and four cards, you may choose at most one card remaining in your hand to bank in your vault, but keep any remaining cards in your hand for your next turn. If and only if you managed to play all five cards, draw five more and take another turn.

4.4 Scoring

At the end of the game, the winner is the player with the highest score, calculated as the sum of the following:

- 1 point per card in your deck, discard pile, hand, and vault
- 3 more points per card in your most numerous suit
- 4 more points per face card (jacks, queens, and kings)
- 5 more points per card in your vault

5 Variations

5.1 The Store

For a shorter game, divide the store into more than three stacks. For a longer game, play until all stacks are empty, not just one. Also, consider allowing a player to buy all store cards simultaneously if an expression may be formed using the store cards that matches the value of the expression formed by the player's cards.

5.2 Clock Arithmetic

Perform all operations modulo 13, or perhaps modulo 12 (and make the kings equivalent to aces).

5.3 Jokers

Add jokers to your deck(s) and treat them as having the value 0; choose suit each time you play one. Alternatively, allow jokers to add or subtract one; to square or take the square root; as factorial; to produce the additive or multiplicative inverse; any other unary operation you can imagine.

5.4 Kings

Give each king a special power. For example: discard K♥ to draw two replacement cards; discard K♣ to put up to two cards from your hand into the vault; discard K♦ to bring a card from the store directly into your hand (available to play in the same turn); discard K♠ to move a card (perhaps only the top card) from an opponent's discard pile into your own discard pile.

5.5 Scoring

In Golomb, make the final score the maximum of the four component scores instead of their sum, or make it the best component score minus the worst component score, or choose one randomly at the beginning of the game, or allow a player to change the criteria mid-game as the reward for a particularly difficult accomplishment.

5.6 Starting Decks

In RPN, since no two cards are alike (unless you have multiple decks) it is impossible for players to start with identical decks. To mitigate the unfairness of a purely random start, use drafting (which is a game skill in itself). To draft, shuffle and deal each player ten cards; choose one card to keep and pass the rest to the left. Repeat this step until you have a complete ten card deck.

If you have only one deck for two players, take all four aces, twos, threes, fours, and kings, and divide them randomly or draft them. For four players, use two decks and the same starting cards. For three players, use two decks and start with all eight aces, twos, threes, and all kings except the kings of spades (K♠). You could also choose which ranks to include in the starting draft randomly (this is analogous to the selection of available cards in Dominion).

Alternatively, if you have enough extra decks that each player can create his or her ideal deck, allow it but perhaps with some constraints such as a cap on the total number of victory points in the deck.

5.7 Final Operation

In RPN, do not use clubs as a mechanism for putting cards in your vault. Instead, whenever you buy a card from the store, you may choose to bank the last card in your expression. For scoring, only count the cards in your vault. This makes the king of spades (K♠) a less desirable card because the only way to claim its 3 victory points is to divide by 13.

5.8 Negative Numbers and Fractions

In RPN, treat red suits (hearts and diamonds) as positive integers and black suits (clubs and spades) as their additive and multiplicative inverses, respectively. In other words, clubs are negative numbers, and spades are unit fractions. Hearts and clubs both add their values; diamonds and spades both multiply their values. Since adding a negative number is the same as subtracting a positive number, and multiplying by a unit fraction is the same as dividing by its denominator, the expression logic is nearly the same as in the base game. The true differences lie in the beginning of the expression and in the target value. If the first card is a 3♣, treat it as a -3. If the first card is a 10♠, treat it as $\frac{1}{10}$, $\frac{1}{5}$, or $\frac{1}{2}$ as you choose. Similarly, to buy a club you will need at least one club in your expression because the expression's value must be a negative number; to buy any spade (other than A♠) you will need at least one spade because the expression's value must be a unit fraction. This variation may work well with the inclusion of a joker in each start deck with the joker serving to take the arithmetic or multiplicative inverse.

5.9 Complex Numbers

I need to think some more about how this would work, but you could treat some suits as real and others as imaginary.

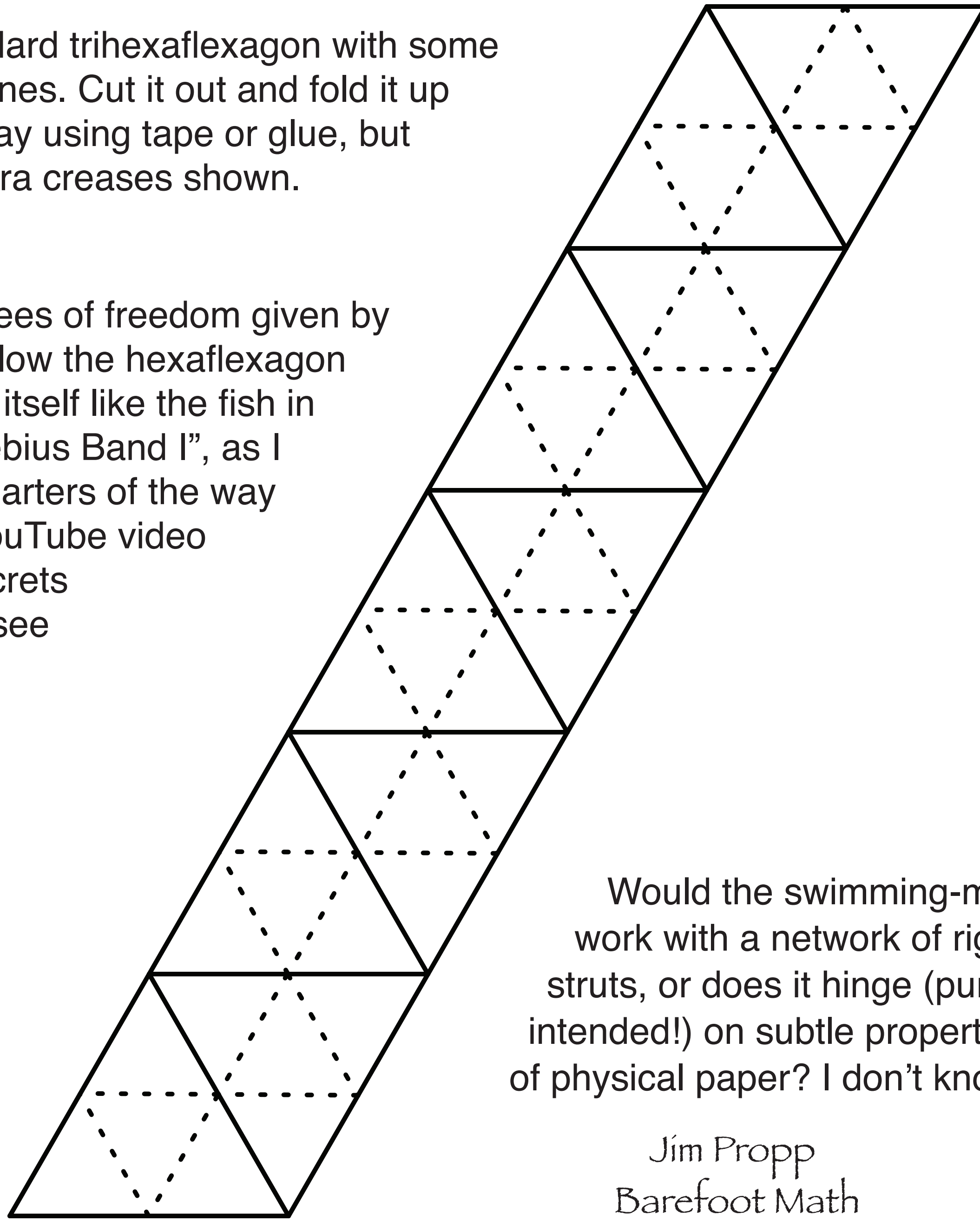
Thanks

I am very grateful to Allison Henle, Steve Stone, and Paul Caginalp for many suggestions; to Miles Henle, Jim Henle, Andreas Santucci, Joe Davis, Rick Bray, and Erica Caginalp for additional playtesting.

A Flexier Hexaflexagon

This is a standard trihexaflexagon with some extra crease lines. Cut it out and fold it up in the usual way using tape or glue, but adding the extra creases shown.

The new degrees of freedom given by the creases allow the hexaflexagon to swim along itself like the fish in Escher's “Moebius Band I”, as I show three-quarters of the way through my YouTube video "Flexagon Secrets Revealed 1" (see 03:20-04:23).



Would the swimming-move work with a network of rigid struts, or does it hinge (pun intended!) on subtle properties of physical paper? I don’t know!

Jim Propp
Barefoot Math
Mathematical Enchantments

An interesting property of Bulgarian solitaire

Tom Roby

May 30, 2018

Abstract

Bulgarian solitaire is a natural discrete dynamical system on the set of integer partitions of a fixed value n. It first arose as a puzzle in the early 1980s, and was popularized by Martin Gardner in one of his *Mathematical Games* columns. Here we focus on an interesting property of this action that came up in joint work with James Propp, namely the *homomesy* ("constant averages over orbits") phenomenon. Showing that Bulgarian solitaire satisfies this property can be considered an extension of the original puzzle.

Divide 15 identical tokens (in practice cards or chips will do) into any number of piles. Take one token from each pile, and make a new pile out of them. For example, if you started with piles of sizes 7,3,3,1,1, your new configuration would have piles 6,5,2,2. (The new pile has size 5 and the singleton piles disappeared.) At the next step, you get 5,4,4,1,1. Continue in this fashion until you can predict the outcome of all future iterations. This is the operation of "**Bulgarian solitaire**", so named by Henrik Eriksson, who notes that it's neither Bulgarian nor a game of solitaire. [Hop12, p. 136]. (How would one win or lose?)

Spoiler Alert: **If you haven't played with this before, I recommend that you try a few rounds first to get a feel for the operation and make your own conjecture.**

The process was presented as a puzzle in Russia in the early 1980s, and Andrei Toom published a solution in *Kvant* [Too81]. Soon after that, it was popularized by Martin Gardner in one of is *Mathematical Games* columns [Gard83]. Brian Hopkins's invaluable article [Hop12] traces the early history and discusses some more recent extensions.

Of course this definition makes sense for any number of tokens. Since we don't care about the order of our piles or the order within our piles, we can formally think of Bulgarian solitaire as a map on *integer partitions* of n, i.e., finite sequences of positive integers in weakly decreasing order $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ satisfying $\lambda_1 + \lambda_2 + \cdots + \lambda_\ell = n$. For example, the partitions of $n = 4$ are

$$\mathcal{P}_4 = \{(1,1,1,1), (2,1,1), (2,2), (3,1), (4)\},$$

and the partitions of $n = 5$ are

$$\mathcal{P}_5 = \{(1,1,1,1,1), (2,1,1,1), (2,2,1), (3,2), (3,1,1), (4,1), (4)\}.$$

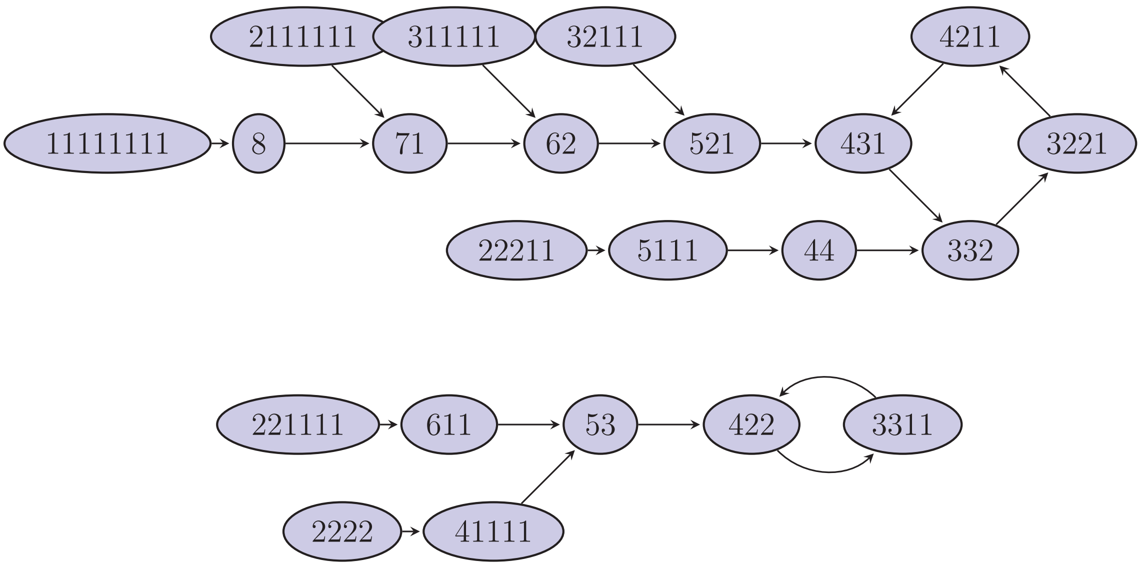
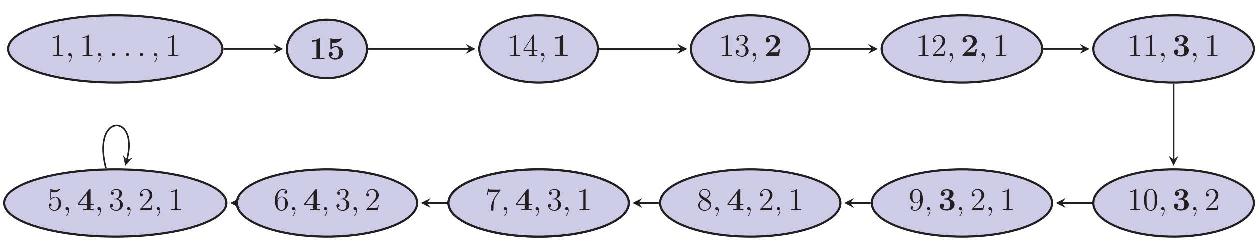


Figure 1: The action of Bulgarian solitaire on all partitions of $n = 8$

In this notation, the map takes $(\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ to the partition whose parts are the *nonzero* elements among $\ell, \lambda_1 - 1, \lambda_2 - 1, \ldots, \lambda_\ell - 1$ (which may need to be re-ordered to be weakly decreasing).

There are 176 partitions of 15. No matter where one starts, one always ends at the "staircase configuration" $(5, 4, 3, 2, 1)$, which is a "fixed point" of this action. The original puzzle was to explain why this happens.

Example 1. Bulgarian solitaire For $n = 15$, one trajectory of Bulgarian solitaire is:



We focus here on a property that came up in joint work with James Propp on the *homomesy* ("constant averages over orbits") phenomenon [PR15, Rob16].

One can show that whenever $n = 1 + 2 + \cdots + k$ is the kth *triangular number*, then any sequence of moves leads eventually to the staircase partition fixed point. A natural question is what happens for more general values of n.

Example 2. Consider Bulgarian solitaire for $n = 8$ as displayed in Figure 1. No matter where we start, we always end up in one of the two "recurrent cycles," namely $(431, 332, 3221, 4211)$ or $(422, 3311)$. The hidden structure here is that the *average* number of parts (piles) is $\dfrac{3+3+4+4}{4} = \dfrac{7}{2} = \dfrac{3+4}{2}$.

The property of a map having constant averages over recurrent cycles (or orbits when the map is interval) was dubbed "homomesy" by James Propp and me [PR15]. A precise definition follows:

Definition 3. Let $\mathcal{S}$ be a *finite* set with a (not necessarily invertible) map $\tau : \mathcal{S} \to \mathcal{S}$ (called a **self-map**). Applying the map iteratively to any $x \in \mathcal{S}$ eventually yields a *recurrent cycle*, and the *recurrent set* is the union of these cycles. (See Figure 1.) We call a statistic $f : \mathcal{S} \to \mathbb{R}$ **homomesic** if the average of f is the same over every recurrent cycle. ($\mathbb{R}$ denotes the real numbers.)

An alternate way to write this definition is to say that the average value of the statistic f as one iterates the map does not depend on the initial starting point, i.e.,

$$\lim_{M\to\infty} \frac{1}{M} \sum_{k=0}^{M-1} f(\tau^k(x)) = c,$$

where c is a constant independent of the starting point $x \in \mathcal{S}$.

In the above example, Bulgarian solitaire acting on partitions of 8, the "number of parts" statistics is ℓ is homomesic with average $\frac{7}{2}$. It is not hard to show that the general situation is as follows, and the proof can be considered an extension of the original puzzle.

Proposition 4 ([Rob16, §2.3])**.** *Let $n = k(k-1)/2 + j$ with $0 \leq j < k$, and consider the action of Bulgarian solitaire on the set of partitions of n. Then the length statistic ℓ which computes the number of parts of λ is homomesic with average $(k-1) + j/k$.*

Note that in Example 2, $n = 8$ corresponds to $k = 4$, $j = 2$, while in the situation that $n = k(k-1)/2$ is a triangular number (so $j = 0$), all paths lead to looping on the shape $\kappa = (k-1, k-2, \ldots, 2, 1)$.

Other statistics homomesic with respect to this action include $f_i(\lambda) := \lambda_i$, the size of the ith largest part of λ, for any $i \geq 1$. For example, when $n = 8$ one sees easily from Figure 1 that f_1, f_2, f_3, and f_4 are homomesic with respective averages $\frac{7}{2}$, $\frac{5}{2}$, $\frac{3}{2}$, and $\frac{1}{2}$.

For more information about the homomesy phenomenon, the reader can consult the original article in which the phenomenon was defined [PR15], or the expository article [Rob16].

References

[Gard83] Martin Gardner, *Mathematical Games: Tasks you cannot help finishing no matter how hard you try to block finishing them*, Scientific American **249** (1983), 12–21.

[Hop12] Brian Hopkins, *30 years of Bulgarian solitaire*, College Math J., **43**, #2, March 2012, 135–140.

[PR15] James Propp and Tom Roby, *Homomesy in products of two chains*, Electron. J. Combin. **22(3)** (2015), #P3.4, http://www.combinatorics.org/ojs/index.php/eljc/article/view/v22i3p4.

[Rob16] Tom Roby, *Dynamical algebraic combinatorics and the homomesy phenomenon* in Andrew Beveridge, et. al., Recent Trends in Combinatorics, IMA Volumes in Math. and its Appl., **159** (2016), 619–652.

[Too81] Andrei Toom, *Problem solution M655*, Kvant (Quantum), **7** (1981), 28–30, http://kvant.mccme.ru/1981/07/p28.htm.

Maverick Solitaire and Three Card Poker

Robert W. Vallin, Department of Mathematics, Lamar University
Beaumont, TX 77710 USA
robert.vallin@lamar.edu

6 March 2018

1 Maverick Solitaire

On January 19, 1958 the American Broadcasting Corporation aired the 17th episode of the first season of *Maverick*, a western starring James Garner. The episode was entitled "Rope of Cards." Garner, portraying gambler Bret Maverick, was the only member of a jury who refused to vote the defendant guilty. During the jury deliberations Maverick claims that it is possible to take 25 cards dealt off the top of a shuffled deck and make five pat poker hands with them. A pat hand is one for which a player would not take extra cards in a game of draw poker. Pat hands are a straight, full house, flush, four of a kind (although one could change out the fifth card in a four of a kind, it will not really improve the strength of the hand, so it is considered pat), and straight flush. In this scene the director created a continuous shot of the twenty-five cards being dealt off and arranged into the five pat hands. It is alleged that there was no set-up involved–it truly was a randomized deck. This game, taking 25 cards and rearranging them, became known as Maverick Solitaire and within days of the episode airing nearly every store in the United States was sold out of playing cards [2].

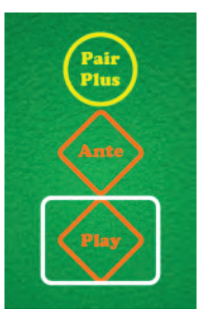
The bet that one can make five pat hands out of 25 random cards is what is referred to as a "proposition bet," where the odds seem to be stacked against the proposer (the person forming the hands) but are actually significantly in that person's favor. Computer simulations say that the proposer will win 98% of the time.

Martin Gardner wrote about this in his book *Mathematical Magic Show* [1]. There he presented the following example of a deal which did could not be made into five pat hands[1].

[1]The key to this is the presence of the four of hearts.

2 Three Card Poker

Three-Card Poker is a casino game. Sometimes derisively called a Midway Game or Carnival Game because the odds of walking away with a profit are significantly against the player. The typical Three-Card Poker table has room for six players to sit, where they are not playing against each other like in a normal poker game. There are two games denoted by two different betting spots: Pairs Plus (where there is no opponent) and Ante-Play against the dealer (see the pictures below). In this note, we are looking at the Pairs Plus game.



Pair Plus Pays

PAIR	1 to 1
FLUSH	4 to 1
STRAIGHT	6 to 1
THREE OF A KIND	30 to 1
STRAIGHT FLUSH	40 to 1

The Pairs Plus Game is straightforward. A player places a bet on the yellow circle, is dealt three cards, and is paid off according to the odds whenever their hand is a pair or better.

In the Ante-Play game a player puts down a bet (the ante) and is dealt three cards. If the player wants to stay in and face the dealer's hand the player places a matching bet on the Play square. The dealer then reveals his/her hand. If the dealer has a queen or

higher, then he/she "qualifies" and the game is played. Whoever has the higher ranked hand wins. Ante pays 1-1 and Play pays 1-1. If the player has a straight or higher, there is an Ante Bonus paid out, too (a straight pays 1 to 1, three of a kind 4 to 1, straight flush 5 to 1). If the dealer does not qualify, the Play bet is returned to the player and the Ante bet pays 1-1.

Both games may be played with the same hand in one deal.

3 Result

The results of this investigation are very similar to the Maverick Solitaire result: Given any five "hands," that is 15 cards, dealt, one can arrange the cards into five Three-Card Poker hands where each hand wins at the Pairs Plus game. This result was found by the taking the fifteen cards and breaking them up into their constituent suits (i.e., checking for flushes) and rearranging as necessary to make non-flush hands. For example,

> In the case of 9 Clubs, 5 Hearts, and one Spade the Clubs and Hearts make 14 cards. There must be at least one card in common. If there are two in common, make two pair and three flushes. If the overlap is only one, then you must have one card of each rank and so there is a straight not involving the overlap card. Thus one straight, one pair, and three flushes. [4]

Since 9 Hearts, 5 Spades, one Clubs, and one Diamond is the same as 9 Spades, 5 Diamonds, one Club, and one Heart we do not have to check all permutations of the suits. This cuts down on the work, but still there were 52 possibilities to consider. Given that, we do not include every breakdown in this note.

We cannot guarantee four Pair Plus winning hands with 12 cards. If we have the Ace through Jack of Spaces and the King of Hearts, the King cannot be a third card in a pair, cannot be used in a straight, or a flush, or a three of a kind.

There is still some openness in questions involving the game. Many, but not every, time at least one hand pays off in the Ante Bonus play. What percentage of the time can this occur? What percentage of the time can one make 4 winning Pair Plus hands and a Straight Flush paying 40-1?

References

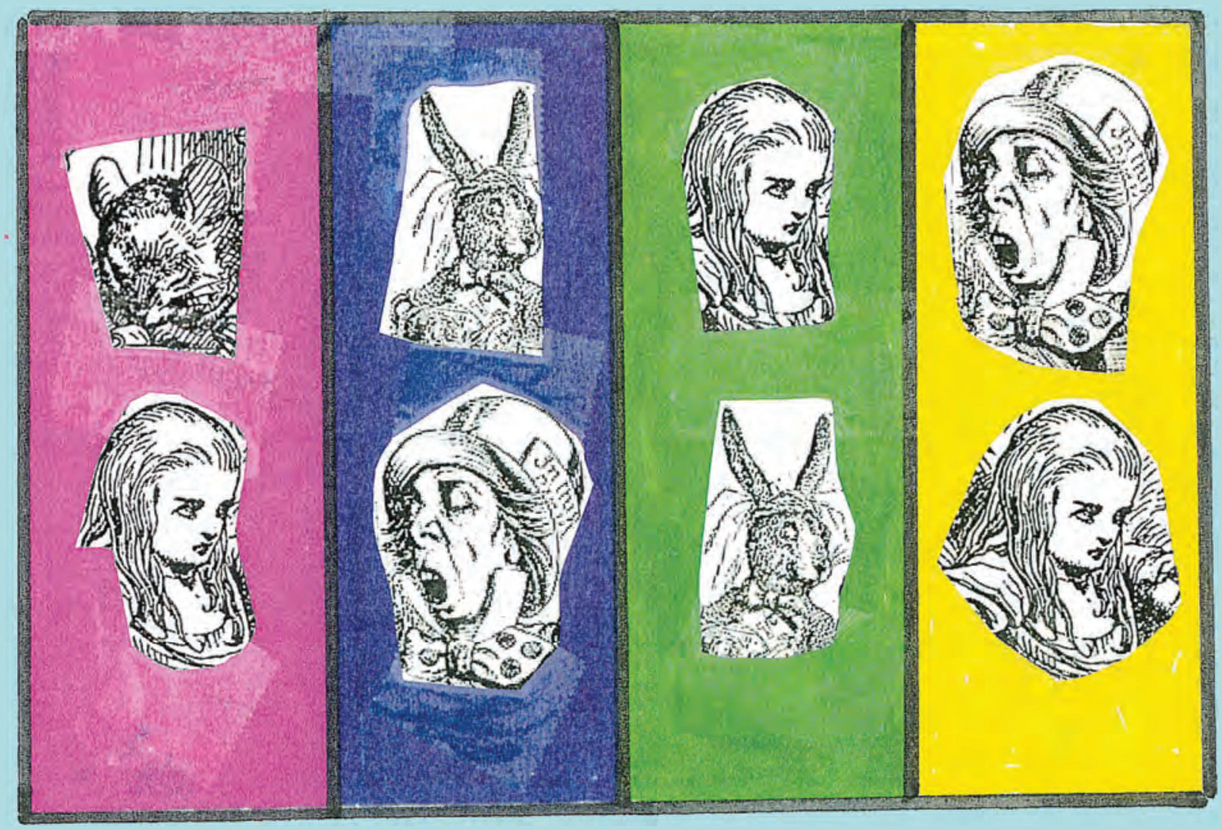
[1] Gardner, Martin, *Mathematical Magic Show*, 94 - 104.

[2] Maverick, "Rope of Cards," International Movie Database, www.imdb.com.

[3] Maverick Solitaire Website, www.solitairelaboratory.com/maverick.html.

[4] Vallin, Robert W., "Three-Card Poker Writeup," (not submitted).

MAGIC



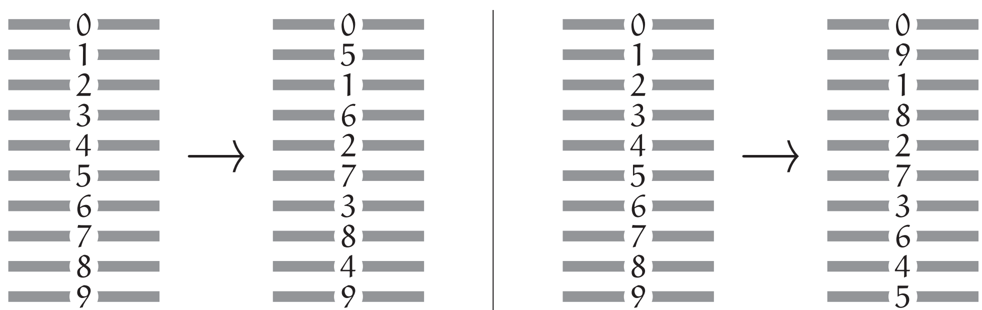
The Tea Party | Stephen Bloom and Jeremiah Farrell | Page 125

A card trick inspired by perfect shuffling

Steve Butler*

Perfect shuffles and horseshoe shuffles

In perfect shuffles we take a deck, split it exactly in half, and then interleave cards from the two sides. This can be done in two different ways and can be distinguished by what happens to the top card (termed **in** and **out**). A variant of this has recently been studied called the horseshoe shuffle which adds one ingredient, namely that before interleaving we reverse the order of one of the halves. The name horseshoe is connected to Smale's horseshoe map, for more information on the mathematics of this shuffle see [1].



Effect of a perfect shuffle | Effect of a horseshoe shuffle

From a mathematical perspective there is a simple connection between the horseshoe shuffle and the perfect shuffle of a deck with twice as many cards. In other words the theory works more or less the same as before. From a performance perspective though there is a strong advantage to working with the horseshoe shuffle. Namely, these involve shuffles that are already named and easier to perform: *milk shuffle* and *Monge shuffle*. So makes for easier tricks to teach people and have them perform. We will give one such example here.

The effect

Take the cards A, 2, . . . , 8 and lay them out on the table so that the audience member can see them and announce "We are going to practice a few basic card shuffling techniques used by magicians."

*Iowa State University, Ames, IA 50010. `butler@iastate.edu`

The cards are now picked up[1] by the performer and they continue, "There are many different shuffles that are used in magic the first one is very simple, we call it dealing down." The performer now deals down one card at a time and picks up the stack. "Of course we can deal down more than one card at a time, since there are eight cards in this stack we can deal down anything which divides eight so we deal down two at a time." The performer deals down two at a time and picks up the stack. "Of we can deal down four at a time." The performer deals down four at a time and picks up the stack. "We can even deal down eight at a time…" The performer places the stack on the table and picks it up again. "…but we usually don't use that one. "Next we have the milk shuffle." The performer demonstrates the milk shuffle (pulling off the top and bottom cards together and placing it on the table and repeating) and picks up the stack. "Finally we have the Monge shuffle." The performer demonstrates the Monge shuffle (moving the cards from one hand to the other, one card at a time alternating above and below). "And the Monge has two variations." The performer demonstrates the other variant (i.e., switch the order of what goes above and below).

At this time the performer hands the deck to the audience member and asks them to now practice the shuffles in any order they want. The performer can have a stunt deck handy to help them remember how to do the various shuffles. Once the audience member is convinced it is well shuffled they are asked to deal the cards face down in the following pattern: left to right, top to bottom (i.e., $\mathsf{a}, \mathsf{b}, \ldots, \mathsf{h}$ as shown below).

a	b	c	d
e	f	g	h

"I am going to try and figure out your cards." The performer now appears to exert some mental energy, but fails. "This works a lot better when I use a marked deck. I need some help, maybe a hint, pick any card you want and turn it over." Suppose the audience member now turns a card over and the performer now sees the following.

a	b	c	d
e	f	g	3

"Ah, the 3, this is helping, the fogs are starting to lift. Since we are doing mathematics let's look for a number that three divides into." The performer points to the third card in the top row and declares, "This card is the 6." The audience member turns it over and sees it is correct.

[1] With a small sleight of hand move to be discussed later

a	b	6	d
e	f	g	3

The performer continues, "I am almost there, just one more hint and I should be able to discern the rest. How about we turn over one of these two cards." The performer points at the cards in positions d and g and suppose the audience member turns over the card in position g to reveal a 7.

a	b	6	d
e	f	7	3

"I see it now! The fog has lifted." The performer now starts pointing at cards and declares what they are, the audience member turns them over discovering that the performer guesses them all correctly!

A	5	6	2
8	4	7	3

How it works

There are several important ingredients to this trick. First and foremost is the fact that this is being done with *eight* cards. What makes this important is that eight is a power of two and it is well known that for powers of two the number of arrangements that happen under perfect shuffles is dramatically smaller than would we expected. For example for eight cards there are $8! = 40320$ different arrangements; but using the shuffles outlined above there will be only 32 different possibilities (a much easier number to handle!). These possibilities follow very specific rules in their structures (see [1]); every shuffle outlined above (and a few more) will preserve this structure. Indeed this is in essence the whole reason why this trick works.

The other important ingredient for us will be the use of binary numbers. We will represent each number as a three-digit binary number (with leading 0's if needed), and we will have 8 correspond to the number 0. So in particular we have the following pairings:

$$\begin{array}{llll} 8 \leftrightarrow 000_{(2)} & A \leftrightarrow 001_{(2)} & 2 \leftrightarrow 010_{(2)} & 3 \leftrightarrow 011_{(2)} \\ 4 \leftrightarrow 100_{(2)} & 5 \leftrightarrow 101_{(2)} & 6 \leftrightarrow 110_{(2)} & 7 \leftrightarrow 111_{(2)} \end{array}$$

The setup

The first part of the trick is to get the cards in the right order. From the above we see that if we left the cards in the order $A, 2, \ldots, 8$ that we would have them in the wrong order in terms of binary. So the one "sleight of hand" is to move the 8 card to the other end of the deck. One easy way to do this is to spread the cards out in order then as you grab the cards you "mistakenly" only grab the first seven cards and then pick up the last card and put it back in the deck (now in the right place).

The first reveal

For all of the 32 possible orderings that can happen the cards are naturally pairing in two ways. One is location, and the other is value. So if we know one card, then we know the location and value of its pair. For location we have the following (essentially notice that this forms a pair of X's).

α	β	γ	δ
β	α	δ	γ

For the values we pair based off of the binary numbers by the following rule: flip the first and last bit. Let's denote this rule as $*-*$, i.e., a $*$ indicates that we flip the corresponding bit and a – indicates that we keep the bit the same. So our four pairs are as follows (of course the performer should embellish as to why these go together):

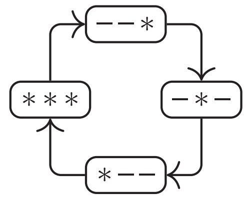
$$\{A = 001_{(2)}, 4 = 100_{(2)}\} \qquad \{2 = 010_{(2)}, 7 = 111_{(2)}\}$$
$$\{3 = 011_{(2)}, 6 = 110_{(2)}\} \qquad \{5 = 101_{(2)}, 8 = 000_{(2)}\}$$

Combining location and value we readily can determine where the other pair is at for whichever card the audience member turns over.

The second reveal

The performer gets the audience member to turn over another card, one which is from a pair that would form a square with the current pair. Using the same technique as from the first reveal we would now have one half of the cards.

To wrap it up we now dip our toes a little more into the structure of how the cards are related. In particular if we look at either of the now uncovered horizontal pairs we would have that there are *four* possible ways that the numbers relate in binary as shown by the boxes in the following diagram.



Whichever way the cards are related in binary (horizontally) follow the arrow to the next box and that will show how the revealed square of shown cards connects to the still hidden set of cards.

This is perhaps best seen by example. So if we go to back to the point in the example performance where the second card has been revealed and apply the first reveal rule we would see in binary that we have the following.

a	b	6	2
e	f	7	3

a	b	110	010
e	f	111	011

At this point we see that (horizontally) the revealed cards are related by switching the leading bit, i.e., $*{-}{-}$. So following the arrow in the diagram we see that the left and right sides are connected by flipping all bits, i.e., $***$. Carrying this out we get the following in binary which we can easily convert back.

001	101	110	010
000	100	111	011

A	5	6	2
8	4	7	3

Notes

This is quite simple and can be learned and taught quickly.

This does take some mild practice to do the reveals mentally. For beginning it is useful to have the following written down: (1) the binary relationships between cards and numbers (8 being 0 is important); (2) the $*{-}*$ pattern of the first reveal; (3) the four patterns and their connections for the second reveal. All this becomes natural with just a bit of practice.

For more information about the mathematics behind this trick look at [1].

References

[1] Steve Butler, Persi Diaconis, and Ron Graham, The mathematics of the flip and horseshoe shuffles, American Mathematical Monthly **123** (2016), 542–556.

THE TEA PARTY

By Stephen Bloom and Jeremiah Farrell

Four coins are used in Tea Party magic. First of all, secretly toss one coin and it will now show up which of either heads or tails stands for Yes while the opposite side will stand for No. The magician will not know which is which.

Now choose, again secretly, one of the four characters at the tea party. The depictions are those drawn by the original artist J. Tenniel.

Trick 1. Secretly choose one of two quirks. "Convivial" and tell the truth to four questions or "Contrary" and lie to all four. Place four coins on the four colors according to your appropriate quirk answering the question "Is your chosen character here?"

Trick 2. Chose now between three quirks: "Convivial", "Contrary" or "Confused". If you choose "Confused" alternately answer true, false, true, false OR false, true, false, true in the order pink, blue, green and yellow.

For either Trick tell the magician on which colors the heads occur. The magician can quickly name your character even not knowing your quirk or whether heads means Yes or No.


AT THE
TEA PARTY


In this Style 10/6

How to determine the character.

On the circular table regard the top and bottom colors to be either pink or green on both. The left and right colors are both either blue or yellow.

Starting at the Mad Hatter on the circle trace either Heads or Tails on the colors. You will always end at the chosen character no matter what quirk was chosen.

CUBIC LOGIC

by Jeremiah Farrell

At the fourth Gathering for Gardner in Atlanta I presented an electronic depiction of a four-dimensional cube on which a magic trick could be performed. The subject was to secretly select a letter from the word ASTEROID and also secretly choose one of two quirks, either CONVIVIAL and always tell the truth or CONTRARY and always lie. After answering each of four questions according to his quirk, the device quickly identified his letter choice. The details are given in my article "Cubist Magic" p. 143 in AK Peters 2002 book *Puzzlers' Tribute, A Feast for the Mind*, edited by David Wolfe and Tom Rodgers.

After my talk Raymond Smullyan, certainly the world's leader in popularizing logic, gently chided me for using CONVIVIAL and CONTRARY as my quirks. He reminded me that mathematicians simply used instead "LIARS" and "TRUTHTELLERS" and these were preferred when talking to them. Of course, I had to agree but when talking to beginning students who have not yet studied logic I still occasionally use CONVIVIAL and CONTRARY.

In fact, I have now added the quirk CONFUSED to my list in which the subject is to alternately lie and tell the truth (starting as he chooses).

Deconstructing Magic Squares

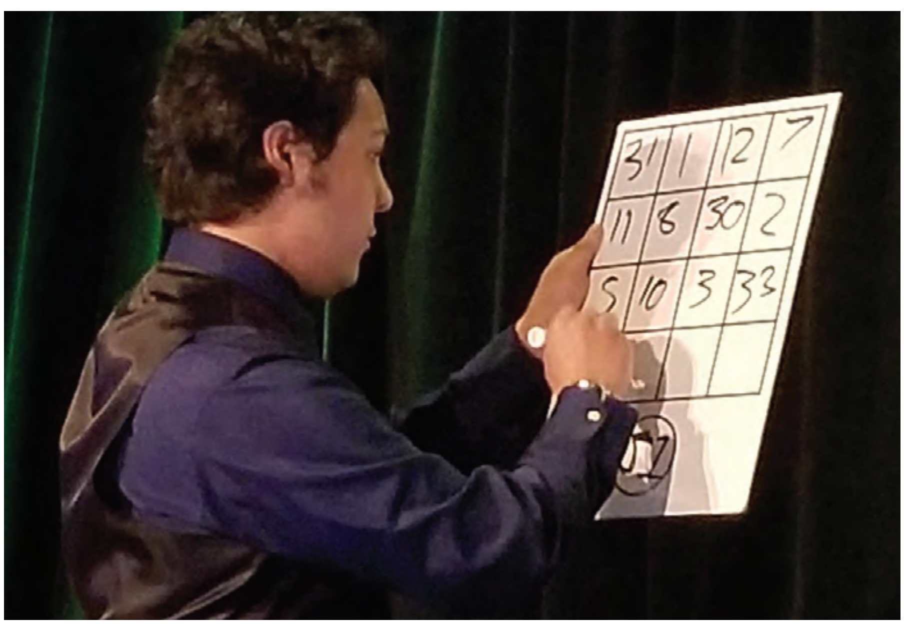
& The MATLAB Magic Show

G4G13 Gift Exchange Paper

Nathaniel Segal

31	1	12	7
11	8	30	2
5	10	3	33
4	32	6	9

51



Magicians have performed magic squares for many years. The general routine involves the magician getting a random 2-digit number (N) that is the magic sum and then the 4x4 grid of numbers then gets filled out quickly to show that all combinations - rows, columns, diagonals, corners, and many others - add up to the given sum.

When I was in the 8th grade, after seeing a magician perform the magic square at a local venue, I was inspired to learn this impressive routine. After months of research and working to solve the square, I figured out that it was a template of numbers 1 - 12 and four variable numbers carefully arranged in one such way:

11	8	N - 21	2
N - 20	1	12	7
4	N - 19	6	9
5	10	3	N - 18

The main secret is that you simply subtract the appropriate numbers from the given sum and plug them in and you get a selected number magic square where nearly all combinations add up. By having 12 set numbers and 4 variable numbers, you are able to make a square with any sum greater than 34 (with no repeated numbers). Because I went about learning this the hard way, I gained a much deeper understanding of the square and the relationships that the numbers had. With the number of performances I studied, I came to notice certain repeating patterns that were just variations of the square above. I then tried to further my research and asked the question if it was possible then to be able to place any number (1 - 12) in any of the 16 squares.

The answer was yes. And what follows is the method for doing just that. I have been performing this informally for some time showing this to small groups or select magicians. This is the first time it is published since I created it nearly 10 years ago. I hope you find this method of interest or potentially some use to you.

The best way to think about the apparent nearly 200 combinations for this new square is to break it down into 5 separate cases.

Case #0: Base Case

I call this Case #0 since it is the simplest scenario since no additional work is required. Let's use the square I showed at the beginning as our standard. Memorize this backwards and forwards. Try to spot patterns or come up with mnemonics to help remember the placements of the numbers and their relationships. I generally recite the square going across the rows, but you should be able to do it from the columns as well.

If they told you to put the number 11 in the top left corner or 9 in the third row and fourth column, then you just need to fill out the general square you have memorized best.

11	8	N - 21	2
N - 20	1	12	7
4	N - 19	6	9
5	10	3	N - 18

Case #1: Rotate Square

Case #1 follows intuitively. This takes very little additional computational effort as all you need to do is rotate the square 90°, or any number of times necessary. So, if 11 if needed to be placed in the top right corner or 9 was placed in the fourth row and second column, then just one rotation will do.

11	8	N-21	2
N-20	1	12	7
4	N-19	6	9
5	10	3	N-18

90° Rotation

5	4	N-20	11
10	N-19	1	8
3	6	12	N-21
N-18	9	7	2

You can see an animation for this at bit.ly/SquareCase1.

Alternatively, you can use a mirrored version of the square, or fill out the square going across rows from a different starting point.

2	N-21	8	11
7	12	1	N-20
9	6	N-19	4
N-18	3	10	5

Mirrored

N-18	3	10	5
9	6	N-19	4
7	12	1	N-20
2	N-21	8	11

Bottom Right Corner Starting Point

Case #2: Swap Rows

Case #2 becomes slightly more challenging. You will simply swap the first and second rows as well as the third and fourth rows. You are performing a symmetrical transformation of the square which retains the rigid nature of the sums. This is what I consider slightly more challenging than Case #1 and less than #3, as I have memorized them going from rows.

11	8	N-21	2
N-20	1	12	7
4	N-19	6	9
5	10	3	N-18

Swap Rows →

N-20	1	12	7
11	8	N-21	2
5	10	3	N-18
4	N-19	6	9

You can see an animation for this at bit.ly/SquareCase2.

Case #3: Swap Columns

Case #3 follows intuitively with the previous case, but this is a swapping of the columns. Simply switch the first two columns with the last two columns.

11	8	N-21	2
N-20	1	12	7
4	N-19	6	9
5	10	3	N-18

Swap Columns →

8	11	2	N-21
1	N-20	7	12
N-19	4	9	6
10	5	N-18	3

You can see an animation for this at bit.ly/SquareCase3.

Case #4: Swap Rows AND Columns

This is by far the hardest case. This is when any of the four corners need to be placed in any of the inside squares or vice versa. Our current methods will not allow us to simply rotate and swap the rows or columns. So, you will need to do both. This takes the most mental energy to map out (without simply memorizing). I carefully place each number to ensure that I have retained the paired relationships between each row and column. This method works because we are still applying symmetrical transformations to the square. With the many combinations, you will get sums from, for example, the four corners as well as the four inside squares. Case #4

is essentially turning the square inside out and the corners become the inside squares and vice versa.

11	8	N-21	2
N-20	1	12	7
4	N-19	6	9
5	10	3	N-18

Swap Columns → AND Rows

1	N-20	7	12
8	11	2	N-21
10	5	N-18	3
N-19	4	9	6

You can see an animation for this at bit.ly/SquareCase4.

These are the 4 basic cases to place any number 1-12 in any of the 16 spots on the magic square. The best thing to do is to practice this many times to get a feel for how you might fill in the square, as there often are more than one possible arrangement for the square that follows all rules set forth here. What I recommend is before going straight into a solve, take a second to think about what the best method to apply might be. Ask if instead of swapping columns you can get a correct solution by starting from a different point. Or, instead of swapping rows, columns, and rotating, if you can use one or two to get the same result. This then becomes a pretty fun game of strategizing and problem solving.

The square on the cover of the paper is a recreation of the one made during my presentation at G4G. To save time but also showcase my method, I incorporated the 4 and 13 aspects by using the number 4 in the 13th spot on the grid (as opposed to asking people to call them out for me) and the number 51 was freely called. You will easily see that this square was a simple application of Case #2 of switching the rows.

In terms of memorization, I came to know what I use as Case #0 as my main point of reference, mainly because I like to remember the numbers in rows and each variable number across rows is larger than the previous one. But, some people might prefer to have the simple calculation of N-20 be the first they do, and so you can figure out whichever square you want to be your point of reference and then apply my method to that square. If you are better at memorization than I am, you can memorize 4 squares (Case #0, #2, #3 and #4) and that will allow you to do the same effect, but only need to apply a rotation, mirroring, or alternative start point (Case #1) to the square. And, if you want to use no mental energy other than recall, then memorizing 16 squares (feel free to figure out which 16) will give you all possibilities for any

number 1-12 in any spot.

From a performance perspective, this is sadly hard to really sell as much more impressive than a standard magic square, at least from an extra effort vs. overall effect perspective. Perhaps you will be able to perform it more effectively than I have been able to, but what I have found is that it impresses magicians and it has about the same effect as the standard square does on the lay audience. If you are thinking of performing this to an audience, I recommend really practicing the different arrangements, and possibly more importantly, practice the strategizing aspect quickly and effectively. This could potentially save you a few seconds from your filling time as well as help you be more confident and less prone to mistakes. I also recommend asking them to place any number 1-9 (instead of 12) in any spot. Saying you can place any number 1-12 seems a bit arbitrary and might make the audience suspicious. It can be easily motivated by noting that you should keep the number to one digit or 1-10 since they are all supposed to add up to the selected sum. I also recommend keeping the sum that is given to a two-digit number. Though it will still work with numbers over 100, the distribution of the variable and set numbers becomes quite skewed and fairly easy to figure out from there.

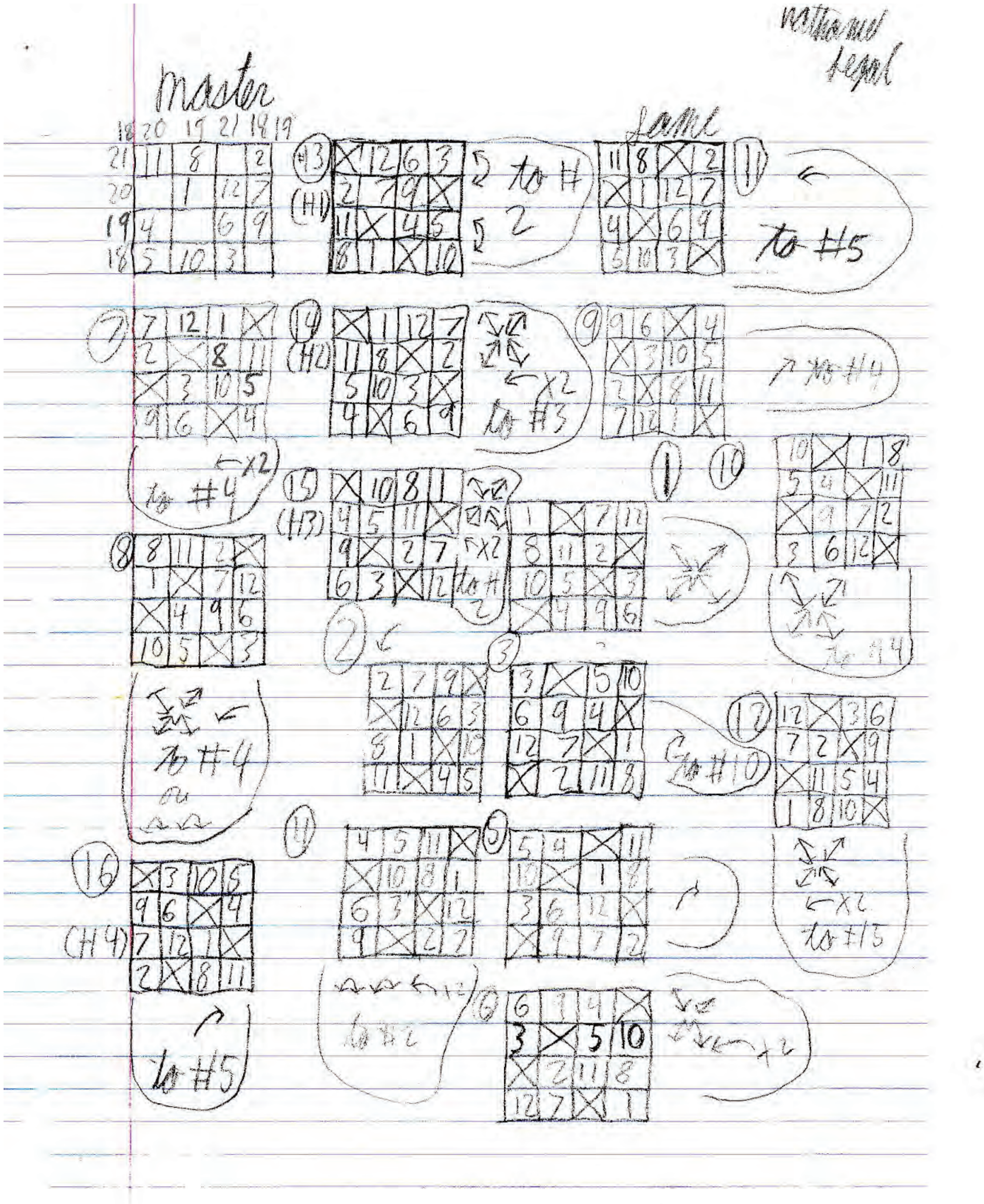
Also, note that with this and the general method, there are two particular combinations that do not add up to the given sum. Looking at the middle two rows, they are the two partitions of the 2x2 squares. One contains two large variable numbers and the other contains none. If performing this, I recommend not calling attention to this and just focus on the sums that actually do add up. It is a bit of a pet peeve of mine when a performer will circle those two combinations along with all other sums to give the appearance of every possible combination working in the hopes that the audience will not catch on or not still checking combinations by then. I believe that some audience members can catch on, and it is best not to call any attention to this discrepancy and let the actual sums speak for themselves.

I am excited to be sharing this with such a special community. I do plan to publish this very soon, but for now you are of a select few who have the method behind my version of the magic square. I do hope you will take the time to learn it and have fun either challenging yourself or sharing this in performance. I would love to hear your thoughts or any questions that you might have. Feel free to email me at magicalnathaniel@gmail.com and I look forward to hearing from you!

<u>Bonus for MATLAB users:</u>

A few years back, I developed a magic show in MATLAB. It is 5 fully interactive (math based) magic tricks that I have added some fun presentation to. All you have to do is run the filename MagicShow and it will give you directions from there. And, one of the tricks is a fully coded version of the magic square that is outlined here. I hope you enjoy!

To download, visit bit.ly/MATLABMagic.



Here is the original sheet made figuring out the different permutations of the square, circa 2008.

How Safe Is It?

Discovering Three Secret Numbers in a Given One

Barney Sperlin, bsperlin@gmail.com

Magic Trick Effect:

A spectator, highly capable in arithmetic OR possessing a calculator, is asked to look at an invisible safe and its invisible combination lock. The lock has numbers from 1 to 6. The magician asks the spectator to turn the dial to one of the 6 numbers, and then turn the dial the other way to some other number, without revealing the two numbers chosen. The spectator is asked to turn the "handle" and open the safe, then put his hand in and drop 1 to 5 imaginary coins inside, close the door and spin the lock.

The magician asks the spectator to subtract the smaller combination lock number from the larger, then multiply the result by 5, then multiply that result by the sum of the 2 combination lock numbers. That result has the number of deposited coins added to it, and the last piece of arithmetic is to multiply the most recent result by 2.

The result is revealed to the magician, who will then use that value as a key to unlock the spectator's mind.

The magician turns the lock with the same 2 numbers secretly chosen by the spectator, announces the numbers after using them, turns the handle to open the safe and reaches in. Pulling his hand out, he opens his hand and it is holding the number of coins deposited, except that they are real.

	2	**3**	**4**	**5**	**6**
1	3	8	15	24	35
2		5	12	21	32
3			7	16	27
4				9	20
5					11

Method:

The formula of the arithmetic above is

$$(\ (\text{comboA} - \text{comboB})*5*(\text{comboA} + \text{comboB}) + \text{coins}\ ) * 2$$

which simplifies to

$$(\text{comboA}^2 - \text{comboB}^2) * 10 + 2 * \text{coins}$$

The magician should memorize the table above. The values in the table are the differences of the squares of the number at the top of the table (one of the combination values) and the values on the left of the table (the other combination value).

For example: 12 (in the table) = $4^2 - 2^2$ and when multiplied by 10 will give 120. If the spectator announces his final result is 126, then the magician considers this as 120 + 6. The magician ignores the zero on the end and, using 12, knows that 4 and 2 were the combination numbers chosen. The 6 is 2 * 3, so there were 3 coins.

If the spectator announces the result as "120", the magician would not think of 12, since there were no coins added if 4 and 2 were the combination numbers. The instructions were to put in from 1

to 5 coins. 120 = 110 + 10 = 11 * 10 + 2 * 5. There were 5 coins, with 6 and 5 as the combination numbers.

This trick is limited to spectators who are unusually good at math and wouldn't be used with general audiences.

Oh, the production of the real coins? The magician has his hand in his pocket, fingering the correct number of coins. The other hand turns the lock and opens the safe door, followed by the hand with the calculated number of coins reaching in, then pulling out to reveal them.

Discussion:

While it is tedious to memorize the chart with the number pair associations, there are some short cuts.

The long diagonal from 3 to 11 contains: a) all odd numbers b) none are skipped c) they are the sum of the matching top and left values and d) the value pair differs by only 1.

The diagonal above it from 8 to 20 has values 4 apart. If you divide one of those numbers by 2 you will get a value which is again the sum of the top and left values on the outside of the chart AND those 2 values are exactly 2 apart in each case.

The diagonal above that contains all odd numbers. If you divide them by 3 you get a number which is the sum of the top and left values, but they are 3 apart.

This may be continued.

Another feature of this trick is that it is only valid for the natural numbers from 1 to 6. Unfortunately, if you use numbers up to 7 then there are values which occur more than once in a table. For example, $7^2 – 5^2 = 5^2 – 1^2$. If you go up to 8, then you also have the problem of $8^2 – 7^2 = 4^2 – 1^2$.

Also, since choosing the same number twice (boxcars or snake eyes with dice, for example) will always result in the value of 0, you can not distinguish between different pairs. That’s why I must make a point of having the spectator choose another number for the second value.

Of course, you can create other scenarios. Invisible dice could be used, or have the spectator mentally choose 2 different boxes from a set of 6 without revealing which 2.

Contact me if you have any questions, suggestions or have another idea for the story line. Try it!

Hamming Code in a Magic Trick

Ricardo Teixeira

January 4, 2018

1 Error Detection and Correction Trick (Hamming Code)

Description: A volunteer thinks of a number and a color, the magician shows cards with several numbers for the volunteer to say whether he sees his number, but he can lie on one card, according to the chosen color. The magician is able to find on which color the volunteer lied, and then tell the number.
Material: Copy and cut cards on Appendix, if you have crayons you could color the cards accordingly.
Preparation: Put the cards in order, they are numbered. Practice how to check parity (see instructions below).
Performance: Gisele, the magician, will read Arthur's mind.

1. Gisele asks Arthur to pick a number between 1 and 15, and one of the colors of the rainbow (red, orange, yellow, green, blue, indigo or violet);
2. Gisele explains that Arthur has to say whether he can see the chosen number on each of the cards she shows;
3. But Gisele also explains that Arthur should tell a lie on the card having the color he chose;
4. For every time he says "yes" for a card, Gisele lays it facing-up, otherwise, if he says "no" she puts the card facing -down;
5. She arranges the cards side-by-side from left to right;
6. Once all seven cards are dealt, she looks at the cards and she can tell which color was chosen, and which number was selected.

Trick: Trick is based on the Hamming Code. The first four cards resemble a binary-digit trick with numbers 1 to 15. If there were no lie allowed, then we'd only need the first four cards. Simply, we would add the top left number on each card that faces up.

However, we are also trying to discover where the lie happened. This is similar to a computer system trying to fix a denigrated data communication. We need more digits (cards).

For every face-up card, you consider a 1 on the sequence. Face-down cards represent 0.

The first four cards will serve to compute the chosen number by adding the top-left number on cards facing up.

The last three cards are the "parity digits".

Check the parity digits (last three cards, see below), if only one of them does not match, then that is the "lie"-card, and the chosen number is the sum of the digits on top left on the first four cards that face up;

If more than one parity digit does not match, then you follow the rule, according to discrepancy:

- Parity 1 and 2: the person lied on the card 3 (yellow), so you flip the card, and the chosen number will be the sum of the numbers on top left on the first four cards who face up (after you "fixed" the sequence);

- Parity 1 and 3: the person lied on the card 2 (orange), so you flip the card, and the chosen number will be the sum of the numbers on top left on the first four cards who face up (after you "fixed" the sequence);

- Parity 2 and 3: the person lied on the card 1 (red), so you flip the card, and the chosen number will be the sum of the numbers on top left on the first four cards who face up (after you "fixed" the sequence);

- Parity 1, 2 and 3: the person lied on the card 4 (green), so you flip the card, and the chosen number will be the sum of the numbers on top left on the first four cards who face up (after you "fixed" the sequence).

Checking Parity Digits

Once you put the cards, consider face-up to be 1, face-down to be 0;

- The first parity digit (the fifth card/digit) needs to make even the sum on digit 2, 3, and 4. For instance, if only one of the digits 2, 3, and 4 is facing-up, then the first parity digit needs to also be 1 (facing-up);

- Second parity digit (sixth card/digit), checks the parity on cards 1, 3, and 4.

- Third parity digit (last card), checks the parity on cards 1, 2 and 4.

Explanation: You are creating a sequence of seven digits 0's and 1's. Because there is a lie, the final sequence will not be any one on the table on "Error Detection" section. With multiple parity checks, we are able to identify the incorrect digit, fix it, and find the correspondent number.
Hint: Practice the error recognition. At first, it may take you a while to figure out the lie. Fix the lie, before telling the chosen number.
Example 1: Suppose that the chosen number is 13, and the chosen color is indigo.

Card 1	Card 2	Card 3	Card 4	Card 5	Card 6	Card 7
Yes	Yes	No	Yes	No	Yes (lie)	Yes
1	1	0	1	0	1	1

Parity 1 (card 5): since between cards 2, 3, and 4, there are two 1's, then it was supposed to be 0. Ok.

Parity 2 (card 6): since between cards 1, 3, and 4, there are two 1's, then it was supposed to be 0. Error.

Parity 3 (card 7): since between cards 1, 2, and 4, there are three 1's, then it was supposed to be 1. Ok.

Since, there is only one parity digit wrong, then that's where the lie is. The chosen number is 8+4+1=13.

Example 2: Suppose that the chosen number is 5, and the chosen color is orange.

Card 1	Card 2	Card 3	Card 4	Card 5	Card 6	Card 7
No	No (lie)	No	Yes	No	Yes	No
0	0	0	1	0	1	0

Parity 1: since between cards 2, 3, and 4, there is only one 1, then it was supposed to be 1. Error.

Parity 2: since between cards 1, 3, and 4, there is only one 1, then it was supposed to be 1. Ok.

Parity 3: since between cards 1, 2, and 4, there is only one 1, then it was supposed to be 1. Error.

Since, parity digits 1 and 3 show discrepancy, the lie is on the second card. Once we fix it, we calculate that the chosen number is 4+1=5.

Extra: How are the cards created?

Now you know how to perform the trick, but how did we place the numbers on the cards in figure 2 so that this trick works? The first four cards give the binary expansion of the chosen number; so these cards correspond to 8, 4, 2, and 1. A number appears on one of these cards if the associated power of two appears in the binary expansion for the number. For instance, the binary expansion for 13 is 13=8+4+1, so 13 appears on the first, second, and fourth cards.

The other three cards are the parity check cards. A number appears on the fifth (blue) card if its binary expansion has an odd number of the following powers of two: 4, 2, and 1. Hence, 9=8+1 and 7=4+2+1 are on the fifth card, but 5=4+1 is not.

The numbers on the sixth and seventh cards are determined similarly. Numbers on sixth card have an odd number of the following powers of 2: 8, 2, and 1. Numbers of fifth cards have an odd numbers of 8, 4, 1.

Further Reading

This magic trick appears on an article by Ricardo Teixeira on the February 2017 edition of Math Horizon, with the title of "Magical Data Restoration." It is a refinement of a trick created by Richard Ehrenborg ("Decoding the Hamming Code," Math Horizons, April 2006) and enhanced by Todd Mateer ("A Magic Trick Based

on the Hamming Code," Math Horizons, November 2013). Ehrenborg's and Mateer's tricks, while essentially the same as this one, require specially designed cards with a clever but little complicated system of tabs to aid the magician. The advantage and disadvantage of using simple rectangular cards is that the trick happens in your mind, instead of physically on the cards.

2 Appendix: Cards

8	9	10	11
12	13	14	15

1: Red Card

4	5	6	7
12	13	14	15

2: Orange Card

2	3	6	7
10	11	14	15

3: Yellow Card

01	3	5	7
9	11	13	15

4: Green Card

01	2	4	7
9	10	12	15

5: Blue Card

01	2	5	6
8	11	12	15

6: Indigo Card

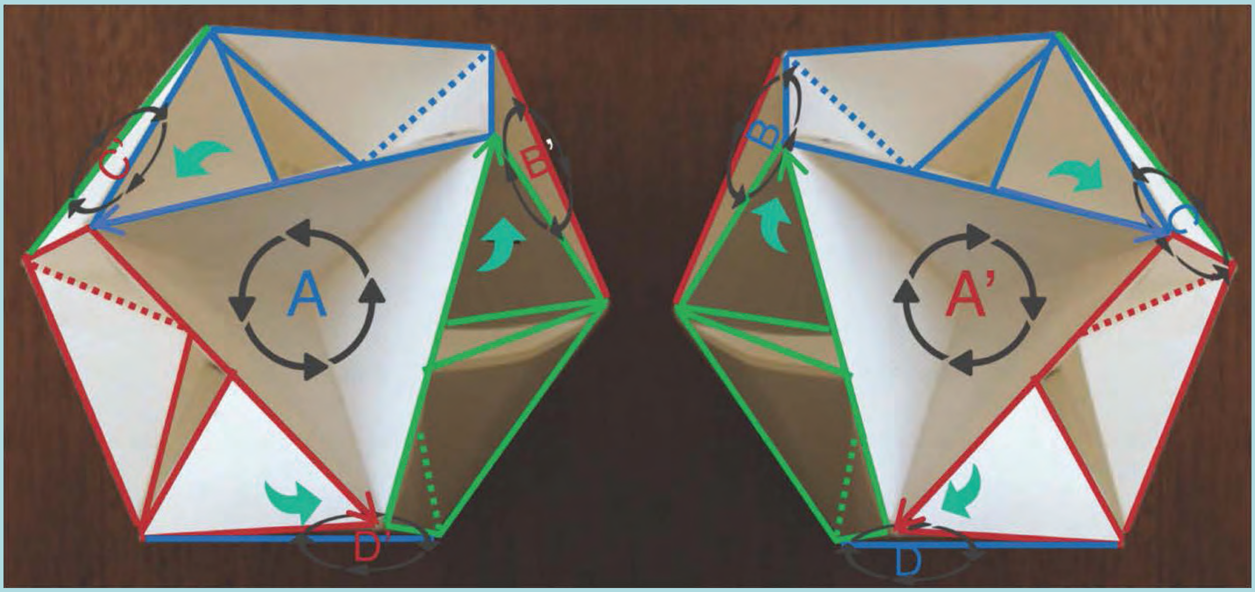
01	3	4	6
8	10	13	15

7: Violet Card

Parity 1 (card 5)	Parity 2 (card 6)	Parity 3 (card 7)
2,3,4	1,3,4	1,2,4

Cheat Sheet

MATH



Chiral Icosahedral Hinge Elastegrity's Geometry of Motion | Eleftherios Pavlides & Peter Fauci | Page 215

Conned Again, Watson! – Gardner's Two Child Problem Re-visited

Anaïs Acree
April 2018
Atlanta, GA

In the book *Conned Again, Watson*, author Colin Bruce takes the reader on adventurous journeys, explaining probability theory through interesting stories. Overall it is insightful and helps explain math by relating it through episodes, but there seems to be some ambiguity in one of his examples. In Chapter 5 titled, “The Case of the Unmarked Graves” he describes a problem closely resembling Gardner’s famous *Two Child Problem*.

The situation (given at the end of this paper) is that a son must know with a better than 50 percent chance that a specific grave is a woman’s burial place before his father will let him dig it open. There are two competing legends, but as the father says, “Now by either legend, the number of male and female skeletons buried here will be equal.” By this statement, there is of course a 50 percent chance that any randomly selected grave will contain the remains of a woman.

To make things more interesting, a shiny woman’s ring is found at a location equally spaced between two graves. All agree that this means that a woman must be buried in at least one of the two graves. The son states, “Father, we know that one of these graves definitely contains a woman’s remains. The other has an even chance of being a man or woman. So, if we dig up one grave, the chances that it contains a female are three in four.”

This assertion argues for an event table similar to Table A where all four events could occur with equal likelihood. Events 1 and 2 are the cases where Grave 1 contains a Female and there is an equal chance that Grave 2 contains a Male or Female. Events 3 and 4 are the cases where Grave 2 contains a Female and there is an equal chance that Grave 1 contains a Male or Female. The probability that Grave 1 is Female given that at least one grave is female is three out of the four events, or 3/4.

Table A

	Grave 1	Grave 2
Event 1	Female	Female
Event 2	Female	Male
Event 3	Female	Female
Event 4	Male	Female

There is, however, another way of looking at this problem as shown in Table B. This event table is populated giving each individual grave an equal chance of being a Male or Female.

Table B

	Grave 1	Grave 2
Event 1	Female	Female
Event 2	Female	Male
Event 3	Male	Female
Event 4	Male	Male

There are three equally likely events with at least one female, (Events 1, 2, and 3). Of these three events, Grave 1 will contain a Female only in Events 1 and 2. This represents a chance of two out of three (not three in four as the son states).

As the story continues, the first grave is opened and contains the remains of a female. They now need to calculate the probability that the second grave also contains a female given that the first contains a female which is analogous to Gardner's "Two Child Problem."

From Table A, this would be two events out of three (Events 1&3 out of Events 1,2,& 3) as described in the book for a probability of 2/3. Table B, however, would give a different answer. Specifically, it would only be one out of two events (Event 1 out of Events 1&2) for a probability of 1/2.

So, which is the right answer? It depends on what can be assumed about the problem and which Event table is correct. Since it is stated that "the number of male and female skeletons buried here will be equal" it could be argued that Table B is correct and the probabilities assigned in the book are incorrect. Another way of thinking about this problem is whether the chance of a ring being present is double if it is between two Female graves rather than if only one of the two graves is Female. In this case, consider Table C showing the Table B possible outcomes with weights assigned for the likelihood of a ring being present.

Table C

	Grave 1	Grave 2	Weight
Event 1 2X	Female	Female	2X
Event 2 1X	Female	Male	1X
Event 3 1X	Male	Female	1X
Event 4 0X	Male	Male	0X

Since Event 1 has two females, it has double the chance, or a weight of 2, for a ring to be present compared to Event 2 or Event 3. Of course, Event 4 has no chance of having a ring since both graves contain Males. Reconstructing Table C to make two separate equally likely events from the doubly likely Event 1, and removing Event 4 (since it would have no rings) essentially reverts back to Table A which gives the results in the book.

To achieve the solution in the book it seems that one must effectively assume that that there is a double chance of a ring being present if it is found between two Female graves compared to between a Male and Female grave. However, this is never explicitly stated in the book, where the finding of the ring is posed as being rather incidental. Specifically, no mention was given to the probability of a ring being present and ironically in the end both graves contained the remains of females but still only one ring was found!

Just like the "Two Child Problem" different people can reach different conclusions depending on how they view this. In my opinion, Table B best describes the equally possible events for this problem. Therefore, instead of the answers of 3/4 and 2/3 as published in the book, I take the position that the answers should be 2/3 and 1/2.

What do you think?! Email me at anaisacree@gmail.com because I would like to know!

"This is it!" said Prendergast excitedly. "This is the ancient burial ground that was used in earlier times. Each of these stones marks the resting place of a king or queen of Arthur's ancient lineage. In a few minutes, we shall know which legend is true. You see, there are two differing stories about how these ancient royals lived and died. Both versions agree that a man must marry to become king, so as to rule with the benefit of both male and female insight, and that king and queen were in due course buried here."

Prendergast eyed us intently as he continued. "Now, one version has it that their marriage was for life, in the modern Christian tradition, and that each was buried when they died of old age or natural causes. But a rival story is darker. It maintains that the king would every seven years take a new young wife, divorcing the old. When the king eventually died, his current wife was immediately killed by beheading, to leave the way clear for the succession. To resolve the matter, we have but to dig up a grave containing a female body and verify that its spine is intact."

I bent down and peered at one or two of the white stones which were the grave-markers. But none seemed to have any writing or other indication upon it. "How can you tell which type of grave is which?" I asked. From behind me there came a harsh laugh: I turned and saw that the Mage had managed, despite his age, to follow us into the bowl.

"Quite so!" he said triumphantly. "I gave my son permission to dig up a grave here—one grave only—on condition that there would be a greater than even chance of finding out the truth. Now by either legend, the number of male and female skeletons buried here will be equal. If you dig up a grave at random, there is an even chance it will turn out to contain a male skeleton, which will tell you nothing. By my edict, you are not allowed to proceed."

Prendergast grimaced, but I could tell by the expression on his face that he saw no way to dispute his father's argument. At his suggestion, we spread out and started to comb the surface of the bowl, though what we hoped to discover was not clear to me. I found myself drawn to a corner where a huge oak tree stood. Its roots had disturbed the ground about it, one great shaft running deep between two of the marker stones. My eye caught something bright in the grass. I bent down and picked up a gold ring, its surface miraculously untarnished, made in the form of the Norse Midgard serpent that lies circling the world with its tail in its mouth. My shout brought the others running.

"It is a woman's ring, a queen's ring," Prendergast shouted. "Show me exactly where you found it."

Alas, I was forced to point to a spot exactly halfway between the two nearest stones.

"You cannot possibly be sure which grave it came from. You do not know which is the female one," said the Mage firmly. Dodgson was about to speak, but Prendergast raised his hand.

"Thank you, Reverend, but I can solve this one for myself. Father, we know that one of these graves definitely contains a woman's remains. The other has an even chance of being man or woman. So if we dig up one grave, the chances that it contains a female skeleton are three in four. By your edict, we may proceed."

The dig took some time, for we proceeded with both caution and reverence. We unearthed the legs, then the pelvis. I was able to identify the pelvis as definitely female, and Prendergast gave a cry of triumph. But as we excavated toward the head, we fell victim to an extraordinary piece of bad luck. A root of the tree had pushed just past the top of the rib cage, and the ground became very wet at that point. Beyond the root there was no further sign of bones. The neck bones and skull, intact or otherwise, were gone.

"I am afraid that the combined work of the tree roots and an underground stream has long since carried that part of the skeleton away, to be scattered and destroyed," I said when it was clear there was no further hope.

Prendergast flung his trowel on the ground in fury. "What an incredible mischance. Really, the gods themselves seem to be against me," he shouted blasphemously. "Father, in the circumstances, may I open the second grave?"

The Mage smiled maliciously. "Of course not," he said. "The ring could obviously have come from this female skeleton, so the chance that the remaining grave is a woman's is again only one in two. I cannot allow you to proceed."

We resumed our seemingly futile search of the bowl. I was devastated that my potentially useful discovery should have led nowhere, and Prendergast must have been feeling far worse. But suddenly there came a cry from the Reverend.

"I have it. Really, I am almost tempted to shout Eureka! The chance that the second grave by the tree root contains a woman is not one-half. It is two-thirds."

The Mage looked at him scornfully. "One-half to two-thirds," he said savagely. "That seems to be your theme song, Reverend, but I am afraid I will not take your word for it. Surely we know nothing about the sex of the second grave."

Dodgson made no reply, but bending to the ground, he picked up a white pebble and a black one. Then he turned to me with a smile.

"Doctor, would you be so good as to lend me your top hat?"

Although somewhat baffled, I gave it to him.

"Let me once again demonstrate the point as a children's game," he said. "I will shuffle these pebbles in my hand and place one at random into the hat. The other I discard without looking at it. So the hat contains a white or a black pebble, with equal probability." We nodded.

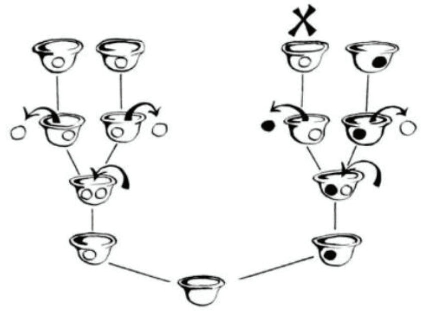
"Now I pick up a second white pebble"—he did so—"and place it in the hat. I toss the pebbles around so I cannot tell which is which.

"If a white pebble denotes a female skeleton and a black a male, I have created a puzzle equivalent to that of the graves by the tree. One is definitely white—that is female. The other is black or white—male or female—with equal probability. Now I take out a stone. I am in luck—it is white." He held up the pebble. "But unfortunately, the stone is mute as to how it met its end." He flung it down. "Now, given that the first stone was white, what is the chance that the second stone is also white?"

He paused. I felt there was something oddly elusive about the problem but was unable to put my finger on it. The others looked equally baffled. At one moment I convinced myself that the probability was only one-third, because we had already used up one female stone, so to speak. Or was it one-half after all?

Dodgson produced a sheet of paper. He pointed to the tree above us.

"The best way to illustrate the possibilities is by drawing a branching tree. I call it my many-worlds tree." He began to draw.



The Many-Worlds Tree

"At the start, we have one version of the world, with a hat that is empty. Now I place in it a stone that may be either white or black, and so we have two versions of the world—two potential worlds in which subsequent events will unfold differently. I add my second white stone, the same in each world, so there are still only two potential realities. Then I take out a stone at random. This may be either the original stone or the second one, so our two worlds fork into four realities."

He counted them across from the top. "In the first reality, I discard the original stone. The second was white, so the remaining stone is white. In the second reality, I discard the second stone, but the original was white. So the remaining stone is again white. In the third reality, I discard the black stone, so that remaining is white. In the fourth reality, I discard the white, and the remaining stone is black."

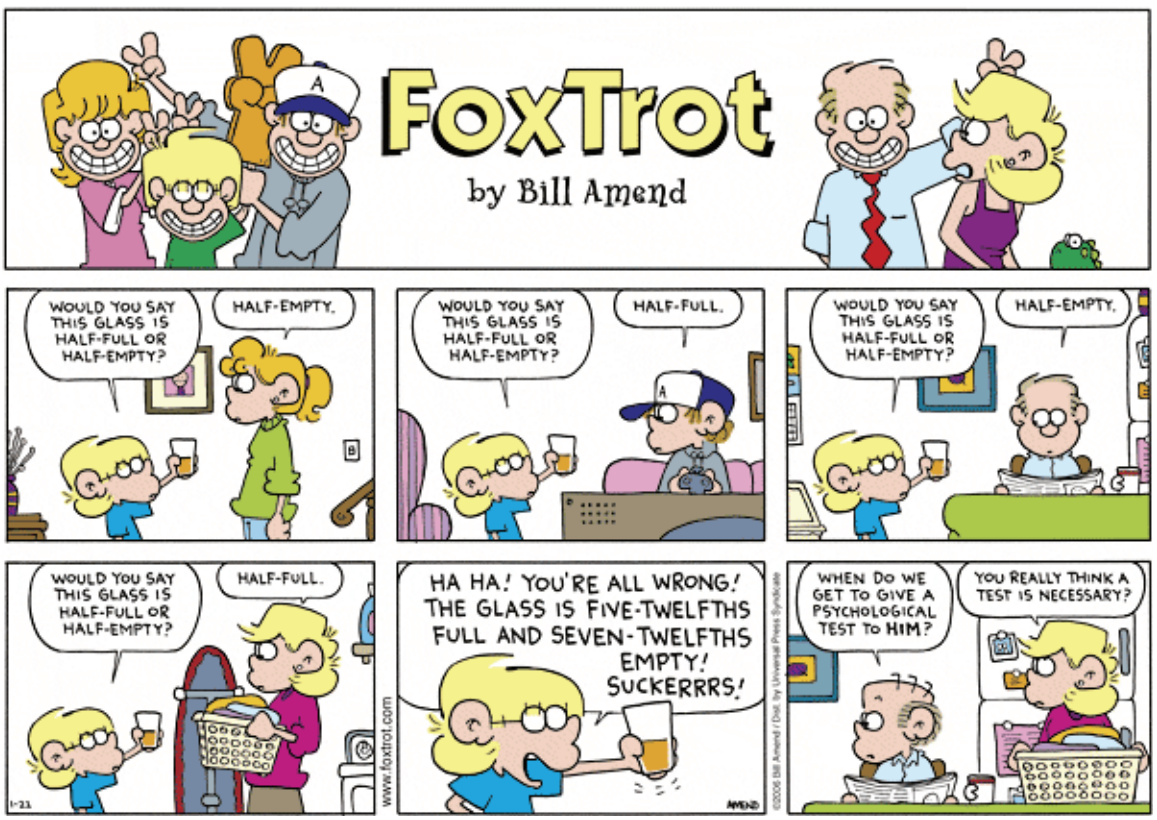
"Then the chances are really three in four that the second stone is white," I exclaimed.

"No, Doctor, because one of the four realities must be crossed out. The first stone I took out was *not* black, so we are definitely not in the third reality. There are three realities we may be inhabiting, each equally likely, and in two of the three the last stone is white. The chances are two in three that the remaining grave contains a female."

The Foxtrot Half-Empty/ Half-Full Problem

Thomas Banchoff, PhD
Professor of Mathematics Emeritus, Brown University
Thomas Cooper, PhD
Professor of Mathematics and Mathematics Education, University of North Georgia.

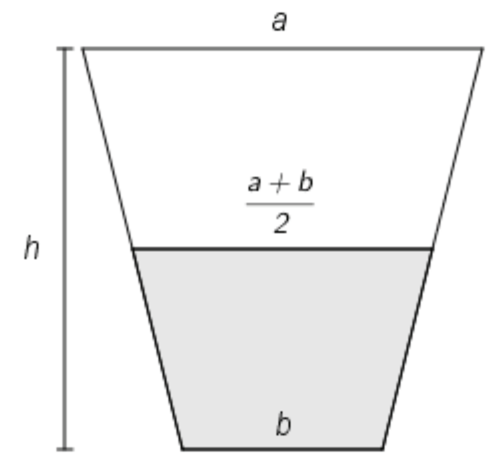
In early 2006, in the Sunday comic strip "Foxtrot", the precocious ten-year-old, Jason shows the members of his family a cup filled up to the halfway level. He asks each one "Is the cup half-empty or half-full?" His brother and mother say "half-full" and his sister and father say "half-empty". He then laughs and says they are all wrong. "The cup is 5/12 full and 7/12 empty! "



Jason's family is not interested in the answer, but we are. There is a good geometry problem suggested here, namely what is the shape of the cup that will yield that precise ratio?

In the semester starting in January 2006, I was a visiting professor at the University of Georgia teaching in the mathematics department and in the School of Education. I taught a geometry course for 25 secondary education majors, and I challenged them to work on the problem and submit their discussions online before the next class.

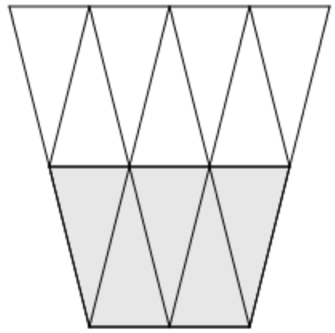
Since the inspiration for the problem was a comic strip, most students started with the two-dimensional problem. The cup was modeled as an isosceles trapezoid with height h, top edge of width a, and bottom edge of width b, shorter than a. The size of the cup was given by the area, with the whole cup having area $h(a + b)/2$. The width of the line halfway up at height $h/2$ was the average $(a+b)/2$.



So the upper region was a trapezoid with height $h/2$, with top edge of width a and bottom edge of width $(a+b)/2$, averaging $[a + (a+b)/2]/2 = (3a + b)/4$ for a total area of $h(3a + b)/8$. Similarly, the area of the lower region was $(h/2)[(a+b)/2 + b]/2 = h(a + 3b)/8$. The ratio of the area of the upper part to the area of the lower part was then $(3a+b)/(a + 3b)$, independent of h. As a check, if $b = 0$ and the cup is a two-dimensional Dixie cup, then the ratio of area of upper part to area of lower part is 3/1, which checks since the lower part is an isosceles triangle and the upper part can be decomposed into three isosceles triangles of the same size.

If b is positive, then we can divide the top and bottom of the ratio by b to get $(3a+b)/(a + 3b) = [3(a/b) + 1)/ ((a/b) + 3)$ so the ratio of upper to lower areas depends only on the ratio $r = a/b$, namely $(3r + 1)/(r + 3)$. To find the unique ratio r giving Jason's ratio (7/12)/(5/12) = (7/5) we need to solve $(3r+1)/(r+3) = 7/5$ so $5(3r+1) = 7(r+3)$ and $15r +5 = 7r + 21$ so $8r = 16$ and $r = 2$. The top edge is twice the width of the bottom edge, which looks like a very good approximation of the ratio of top edge to bottom edge of the cup in the comic strip.

Even though the analytic geometry argument based on the work of several students was quite convincing, even more convincing was a one-diagram geometric "proof by picture". A trapezoid with top edge of width 4 and bottom edge of width 2 can be divided into 12 congruent isosceles triangles with base of width 1, with 7 triangles in the top part and 5 in the bottom part! One student wrote the single word "Walla!" on the bottom of this diagram and when I asked what he meant by that, he said "You know, the French word, Voila!"



Although this kind of proof is quite convincing when the ratio of the top width to the bottom width is a rational number, there is no good diagram of the same sort when that number is irrational.

As it happens, there were three PhD candidates observing the class, and one of them, my G4G13 co-presenter Tom Cooper, now a full professor at the University of North Georgia. He interpreted the Foxtrot problem as three-dimensional and that led up to a more complicated situation. Instead of a trapezoid, he considered the frustum of a right circular cone of height h with radius a for the top circle and radius b for the bottom disc. In the class we had already showed that the volume of such a cup is $(\pi/3)h(a^2 + ab + b^2)$. A horizontal slice at height $h/2$ would have radius $(a+b)/2$, so the volume above that slice would be $(\pi/3)(h/2)[a^2 + a(a+b)/2 + ((a+b)/2)^2]$ and the bottom below the slice would have volume $(\pi/3)(h/2)[((a+b)/2)^2 + ((a+b)/2)b + b^2]$.

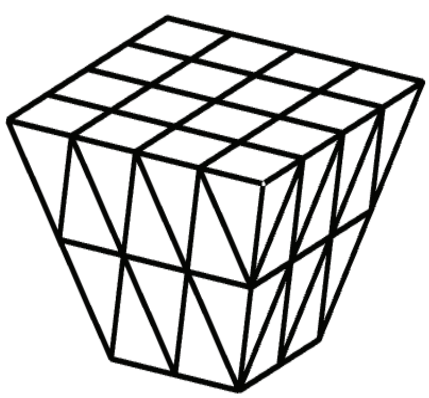
As in the two-dimensional case, the ratio of the volume of the upper region to the volume of the lower region is independent of h and only depends on the ratio $r = a/b$ of a to b. The formula for this ratio is more complicated in the three-dimensional case, involving quadratic factors:

$$\frac{a^2 + a \cdot \frac{a+b}{2} + \left(\frac{a+b}{2}\right)^2}{\left(\frac{a+b}{2}\right)^2 + \frac{a+b}{2} \cdot b + b^2} = \frac{r^2 + \frac{r(r+1)}{2} + \frac{(1+r)^2}{4}}{\frac{(1+r)^2}{4} + \frac{r+1}{2} + 1}$$

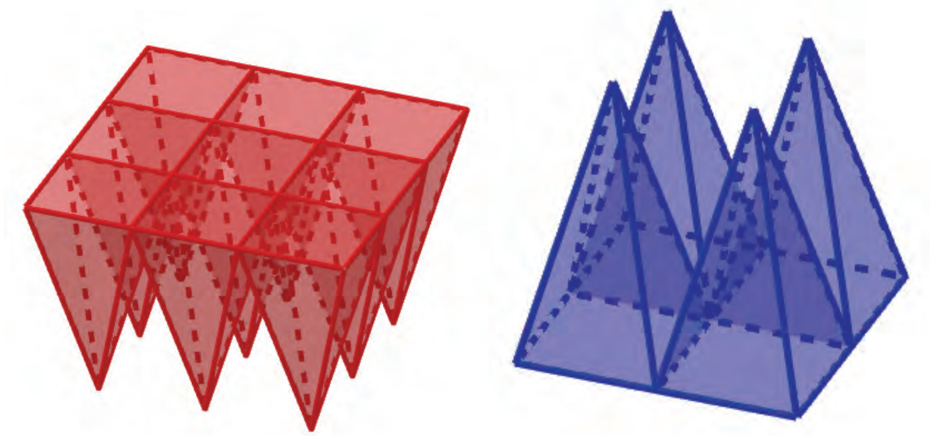
If $r = a/b = 2$, then the ratio of the top part to the bottom part is $(4 + 3 + 9/4)/(9/4 + 3/2 + 1) = (37/4)/(19/4) = 37/19$, not such a simple calculation as in the two-dimensional case. So if Jason lived in a three-dimensional world, he could have said, “The cup is 19/56 full and 37/56 empty!”

That solves the problem for a surface of revolution of an isosceles trapezoid, but Jason’s planar cup has different analogs in the third dimension. For example, a truncated inverted “Egyptian” pyramid with square top of side length a and square bottom of side length b. Analogous to the volume of the truncated cone, the truncated pyramid has volume $V = (1/3)h(a^2 + ab + b^2)$.

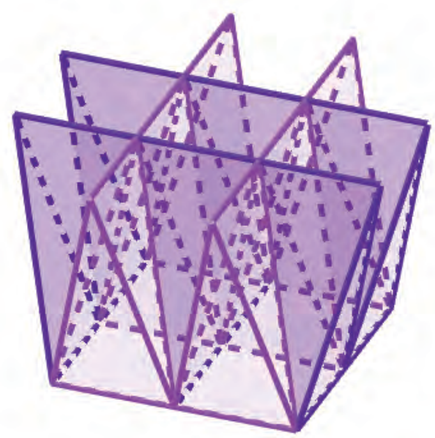
For the case of $a = 4$ and $b = 2$, there is an analogous demonstration similar to the “Viola” case using pyramids and tetrahedra. The halfway slice is a square with side length 3.



We can divide the lower region into four square pyramids pointing up and nine congruent pyramids pointing down (producing 13 congruent pyramids for G4G13). Those sets of pyramids correspond to the terms a^2 and b^2 in the formula for the volume of a frustum of a square pyramid.



In addition, there are 12 tetrahedra corresponding the *ab* term in the formula for the volume of a frustum.



The truncated cones and pyramids are only two kinds of generalizations of Jason's two-dimensional cup. There are other considerations in dimension three, or even higher! We would like to thank Jason and his creator Bill Amend for inspiring our geometric excursion.

Why Do the Unit Quaternions Double-Cover the Space of Rotations?

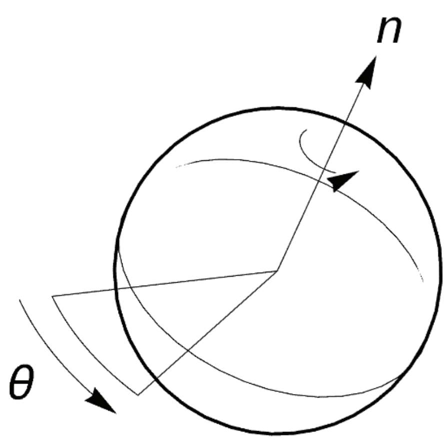
Neil Bickford

1. Computing with Quaternions

The unit quaternion

$$\boldsymbol{q} = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ x\sin\left(\frac{\theta}{2}\right) \\ y\sin\left(\frac{\theta}{2}\right) \\ z\sin\left(\frac{\theta}{2}\right) \end{pmatrix}$$

represents a counterclockwise rotation by the angle θ around the normalized axis $\boldsymbol{n} = (x, y, z)^T$.



We'll sometimes use q_w, q_x, q_y, and q_z to refer to the four components of a quaternion.

We can compose quaternions in the same way we can compose rotations: the product **r** of quaternions **p** and **q**

$$\boldsymbol{r} = \boldsymbol{pq}$$

represents the rotation given by performing $\boldsymbol{q}$, then by performing $\boldsymbol{p}$. For instance, if **q** is a rotation by the angle θ around the axis $(x, y, z)^T$, then the product of **q** with itself is twice the rotation:

$$\boldsymbol{q}^2 = \left(\cos\left(\frac{2\theta}{2}\right), x\sin\left(\frac{2\theta}{2}\right), y\sin\left(\frac{2\theta}{2}\right), z\sin\left(\frac{2\theta}{2}\right)\right)^T.$$

Similarly,

$$\boldsymbol{q}^3 = \left(\cos\left(\frac{3\theta}{2}\right), x\sin\left(\frac{3\theta}{2}\right), y\sin\left(\frac{3\theta}{2}\right), z\sin\left(\frac{3\theta}{2}\right)\right)^T$$

and so on. The inverse of a unit quaternion[1] is given by reversing its rotation:

$$\boldsymbol{q}^{-1} = \left(\cos\left(-\frac{\theta}{2}\right), x\sin\left(-\frac{\theta}{2}\right), y\sin\left(-\frac{\theta}{2}\right), z\sin\left(-\frac{\theta}{2}\right)\right)^T.$$

If we ever need to, we can write out the result of performing $\boldsymbol{q}$, then $\boldsymbol{p}$ (like finding the rotation that corresponds to a product of two other rotations), as another quaternion:

$$\boldsymbol{pq} = \begin{pmatrix} p_w q_w - p_x q_x - p_y q_y - p_z q_z \\ p_x q_w + p_w q_x - p_z q_y + p_y q_z \\ p_y q_w + p_z q_x + p_w q_y - p_x q_z \\ p_z q_w - p_y q_x + p_x q_y + p_w q_z \end{pmatrix}$$

This gives us a way to express the product of two quaternions. Note that we have to be careful about the order in which we apply quaternions (and rotations); for instance, a 90° rotation around the *x* axis followed by a 90° rotation around the *y* axis produces a different result than a 90° rotation around the *y* axis followed by a 90° rotation around the x axis.

We can also express this as a matrix-vector product, which is useful for computer implementation: if $\boldsymbol{r} = \boldsymbol{pq}$, then

$$\begin{pmatrix} r_w \\ r_x \\ r_y \\ r_z \end{pmatrix} = \begin{pmatrix} p_w & -p_x & -p_y & -p_z \\ p_x & p_w & -p_z & p_y \\ p_y & p_z & p_w & -p_x \\ p_z & -p_y & p_x & p_w \end{pmatrix} \begin{pmatrix} q_w \\ q_x \\ q_y \\ q_z \end{pmatrix}$$

We can even express quaternions themselves as 4x4 matrices and have all of the normal notation carry over:

$$\begin{pmatrix} r_w & -r_x & -r_y & -r_z \\ r_x & r_w & -r_z & r_y \\ r_y & r_z & r_w & -r_x \\ r_z & -r_y & r_x & r_w \end{pmatrix} = \begin{pmatrix} p_w & -p_x & -p_y & -p_z \\ p_x & p_w & -p_z & p_y \\ p_y & p_z & p_w & -p_x \\ p_z & -p_y & p_x & p_w \end{pmatrix} \begin{pmatrix} q_w & -q_x & -q_y & -q_z \\ q_x & q_w & -q_z & q_y \\ q_y & q_z & q_w & -q_x \\ q_z & -q_y & q_x & q_w \end{pmatrix}$$

(We'll show how to derive this in section 5.)

Alternatively, here's an easy way to remember the product of two quaternions: Imagine we extend the real numbers with three symbols *i, j,* and *k* (in the same way that we can extend the real line to get the complex numbers) with the properties that

$$i^2 = -1, j^2 = -1, k^2 = -1, \text{and } ijk = -1.$$

[1] There are such things as non-unit quaternions, which we won't talk about in this article outside of this footnote. They can be thought of as combinations of a rotation and a scale by the length of the quaternion, the square root of the *norm* given by $N(\boldsymbol{q}) = |q|^2 = q_w^2 + q_x^2 + q_y^2 + q_z^2$. The unit quaternions are those quaternions with norm 1: $q_w^2 + q_x^2 + q_y^2 + q_z^2 = 1$. For the inverse of a general quaternion, divide the coefficients above by $N(\boldsymbol{q})$.

This norm happens to satisfy $N(\boldsymbol{pq}) = N(\boldsymbol{p})N(\boldsymbol{q})$ for any two quaternions $\boldsymbol{p}$ and $\boldsymbol{q}$, which gives a quick way to derive Euler's four-square identity. As it turns out, the requirement for such a norm to exist is the main reason why normed division algebras over the reals are only possible in dimensions 1, 2, 4, and 8.

From these properties, we can derive the product of any two basis elements: $ij = k, ji = -k, jk = i, kj = -i, ki = j$, and $ki = j$. We can then write the quaternion $\boldsymbol{q} = (q_w, q_x, q_y, q_z)$ as

$$q_w + q_x i + q_y j + q_z k$$

and have all the normal multiplication work:

$$\begin{aligned}
&(p_w + p_x i + p_y j + p_z k)(q_w + q_x i + q_y j + q_z k) \\
&= (p_w q_w - p_x q_x - p_y q_y - p_z q_z) + (p_w q_x + p_x q_w)i + (p_w q_y + p_y q_w)j + (p_w q_z + p_z q_w)k + p_x q_y ij \\
&\qquad + p_x q_z ik + p_y q_x ji + p_y q_z jk + p_z q_x ki + p_z q_y kj \\
&= (p_w q_w - p_x q_x - p_y q_y - p_z q_z) + (p_x q_w + p_w q_x - p_z q_y + p_y q_z)i \\
&\qquad + (p_y q_w + p_z q_x + p_w q_y - p_x q_z)j + (p_z q_w - p_y q_x + p_x q_y + p_w q_z)k.
\end{aligned}$$

With the equations above, quaternions give us a fast and efficient way to store rotations, express the composition of rotations, and, importantly, to *smoothly blend* between rotations (which we'll cover in section 5.)

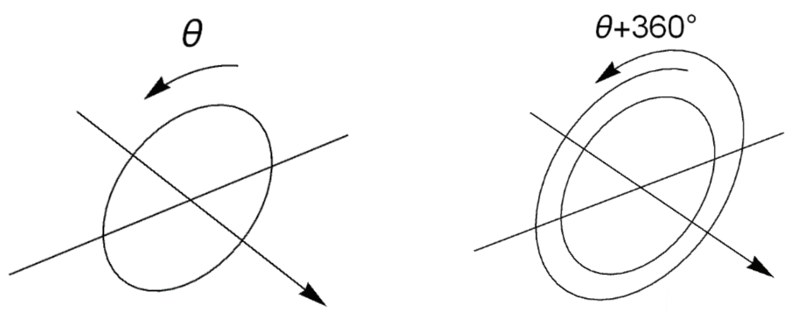
Now, here's something interesting: Consider the rotation around an axis $(n_x, n_y, n_z)^T$ by an angle θ, represented by a quaternion:

$$\boldsymbol{q} = \left(\cos\left(\frac{\theta}{2}\right), n_x \sin\left(\frac{\theta}{2}\right), n_y \sin\left(\frac{\theta}{2}\right), n_z \sin\left(\frac{\theta}{2}\right)\right)^T.$$

If we rotate around this axis by an additional 360°, we get

$$\boldsymbol{q}' = \left(-\cos\left(\frac{\theta}{2}\right), -n_x \sin\left(\frac{\theta}{2}\right), -n_y \sin\left(\frac{\theta}{2}\right), -n_z \sin\left(\frac{\theta}{2}\right)\right)^T$$

This is a different quaternion, but *it represents the same rotation*; we've just rotated an extra 360°.



Rotating an additional 360°, for a total of 720°, brings us back to the first quaternion.

In fact, we see the following: Every rotation – an angle around some axis – is represented not by one, but by *two* quaternions. In this way, we say that the unit quaternions *double-cover* the space of rotations. But why do quaternions *have* to double-cover rotations? Could we – say – just do something like replacing $\frac{\theta}{2}$ by θ?[2]

[2] Answer: No, because then all 180° rotations would be represented by the same four-vector, $(-1, 0, 0, 0)^T$.

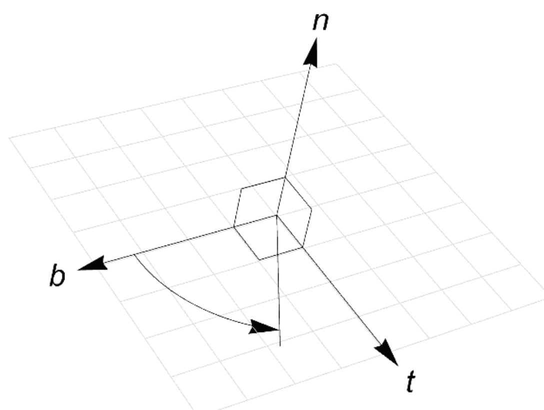
This article is about this phenomenon of double-covering. In short: When we're talking about interpolating between paths in the space of rotations, it actually matters how many times our rotation has completed a full circle.

In a wider mathematical frame, it turns out that the space of unit quaternions is actually slightly nicer than the space of rotations: the space of quaternions maps nicely to a sphere in four dimensions, while the space of rotations isn't simply connected. Put another way, the unit quaternions cover the space of rotations twice, because they cannot cover the space of rotations once and also provide a way to interpolate between paths of rotations. By working with quaternions, we get to work with points on a sphere, instead of points on a real projective plane.

2. An Orthogonal Basis Problem

Here's a system of representing rotations that doesn't work.

Suppose we want to rotate a point around an axis. One way to do this might be to extend the axis into a full coordinate frame, by somehow finding two additional vectors which meet at right angles to the axis and to each other.



Once we have such a coordinate frame[3], we can map our point to our coordinate system, rotate around the axis using a two-dimensional rotation in the plane of our other two vectors, and then transform back into the original coordinate system.[4]

$$R = \begin{pmatrix} | & | & | \\ \boldsymbol{n} & \boldsymbol{b} & \boldsymbol{t} \\ | & | & | \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} | & | & | \\ \boldsymbol{n} & \boldsymbol{b} & \boldsymbol{t} \\ | & | & | \end{pmatrix}^{-1}$$

Ideally, we'd also like our two additional vectors to vary smoothly as we adjust our axis. Put another way, we'd like to find some continuous function which takes as input a normalized vector and

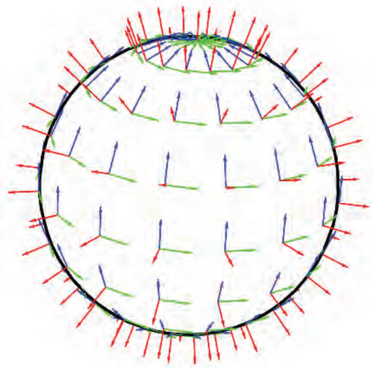
[3] Technical note: With a fixed handedness.

[4] If ***n, b***, and ***t*** are normalized, then this is slightly easier; since the basis transform matrix is orthogonal, we have

$$\begin{pmatrix} | & | & | \\ \boldsymbol{n} & \boldsymbol{b} & \boldsymbol{t} \\ | & | & | \end{pmatrix}^{-1} = \begin{pmatrix} | & | & | \\ \boldsymbol{n} & \boldsymbol{b} & \boldsymbol{t} \\ | & | & | \end{pmatrix}^{T}.$$

outputs the rest of the coordinate system. (The reason we want this function to be continuous is because we'd like to be able to nicely interpolate between rotations; otherwise, although the results might be OK, our internal model of the system would suddenly change its state as we passed over the discontinuity.)

Unfortunately, this is impossible: no matter how we try to construct such a function, we'll always have a discontinuity somewhere in the space of unit vectors. The easiest way to see this is through the Hairy Ball Theorem: if we had such a function, then we'd be able to place a coordinate frame at each point of a sphere, like this (excepting the north and south poles in this illustration):



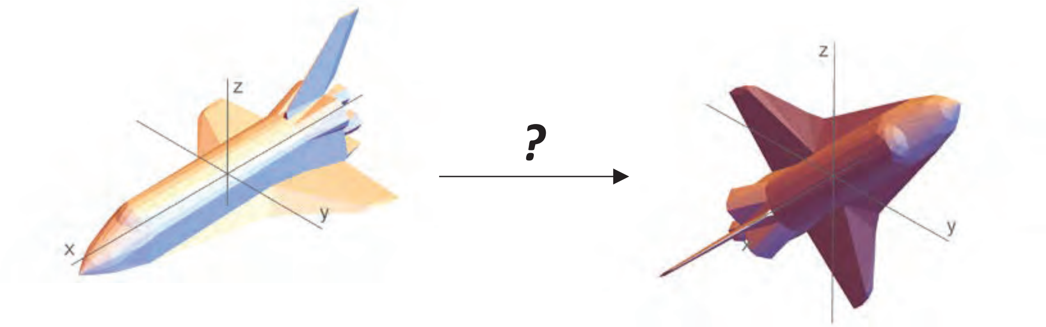
If we could do such a thing, then the green or blue vectors would form a continuous vector field on the surface of the sphere. But this would produce a smooth combing of the sphere, which is impossible by the Hairy Ball Theorem.

3. Euler Angles

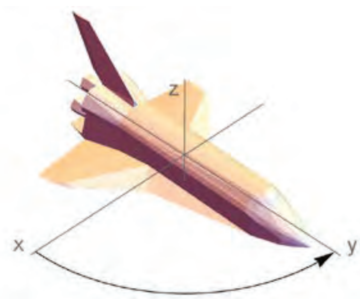
If you've heard of Euler angles before, these singular points that arise when trying to comb the sphere might remind you of *gimbal lock*, a problem that arises in using this system to represent rotations – where at particular choices of axis, you lose a degree of freedom and the entire system freezes up until you rotate it out or through of this particular range of axes. If you haven't heard of Euler angles before, let's rewind a bit.

Euler angles give us a way to represent orientations as a unique product of a yaw, a pitch, and a roll, as follows:

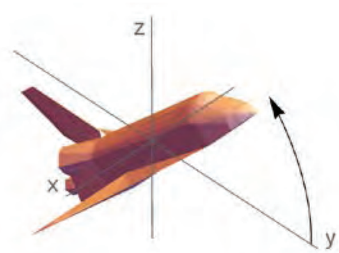
Suppose we have an object which we want to rotate to a particular orientation, which we express (as with quaternions) as an axis and an angle around that axis.



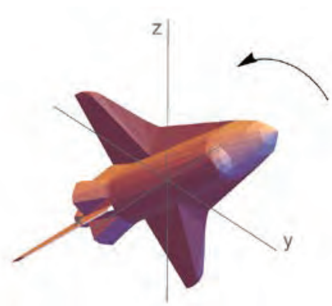
First, we match the lateral bearing (yaw) of the object:



Then, we pitch the object up or down to match the new axis:

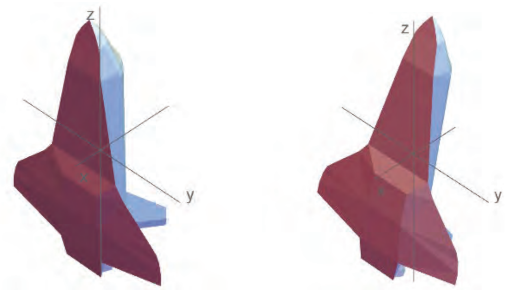


Finally, we rotate (roll) the object around this axis to match the full orientation.



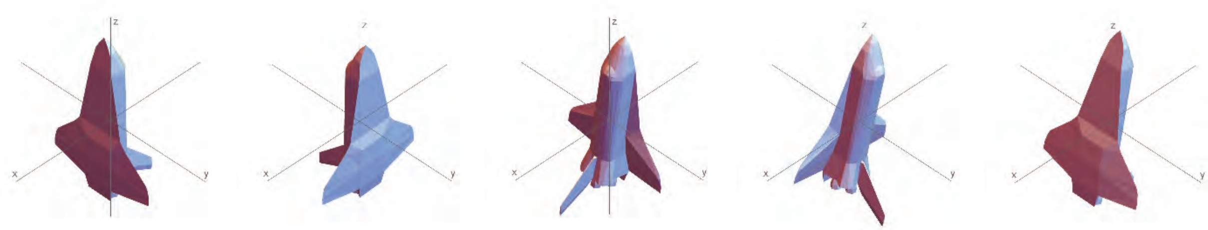
If we limit the yaw, pitch, and roll to some range of values (e.g. $0 \le yaw, roll < 360°$ and $-90° < pitch < 90°$), these three numbers can then be used to uniquely represent any orientation outside of gimbal lock. (We can also uniquely specify a rotation instead of an orientation by listing the yaw, pitch, and roll that a model undergoes as a result of that rotation.)

However, we have a problem: Regardless of our choice of Euler angle system[5], we'll always be able to find some axis near which slightly different orientations lead to wildly different Euler angles. For instance, consider these two orientations with regards to the above system:



In the above system, the first of these two orientations can be given by a simple 90° pitch upwards. For the second, we need to turn the object 180°, pitch it upwards just less than 90°, and then roll it another 180°.

In particular, this means that our mapping in reverse from orientations to Euler angles is discontinuous – and that as a result, in order to interpolate between two orientations using Euler angles, we might have to take a longer route than necessary:



Gimbal lock has been the cause of a variety of problems in real-world systems; for more information on the ill effects of gimbal lock, see [Hanson, pg. 19-27].

4. The Connectedness of Rotations

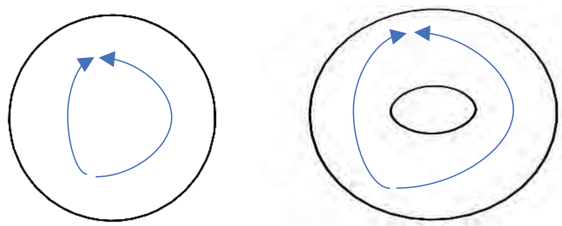
Euler angles, in fact, have problems for reasons beyond the Hairy Ball Theorem – their geometry is that of a four-dimensional torus (which doesn't correspond to the geometry of the space of rotations), and the fact that they have three parameters also prevents them from smoothly representing the space of rotations (if we consider the sets of points we get as we vary the angle of rotation from 0° to 360°, we

[5] We've described the ZYX system of Euler angles above, where we rotate around the Z, Y, and X axes in our local coordinate frame in sequence. (For instance, at the second step, we rotated around the object's local Y axis.) We can also describe Euler angles as a triplet of rotations around each of three axes, at least one different from the rest. We'll always have gimbal lock somewhere, regardless of the Euler angle system we choose; however, in some cases (e.g. ships), we might be able to choose a frame such that we should never rotate to an orientation that would cause gimbal lock.

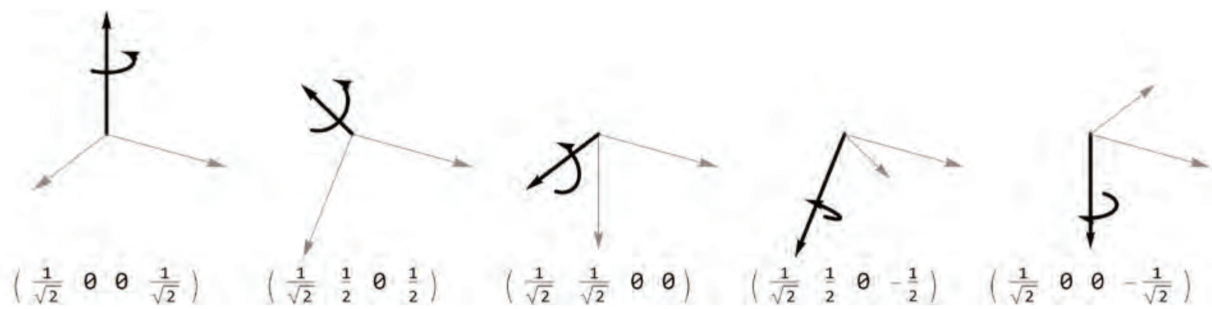
can see that the sets of spheres we get must at some point "turn back" on themselves, which prevents us from having a nice mapping from a region in $\mathbb{R}^3$ to the space of rotations.)

Even if we uniquely parameterized rotations as a pair of an axis and an angle (with some constraints, so that we express each rotation exactly once), we'd still run into problems when talking about interpolation – specifically, when talking about interpolation between paths of rotations.

Consider a space, and draw two continuous paths through the space which begin and end at the same two points. We say that this space is *simply connected* if, no matter which two points or paths we choose, we can always continuously transform the first path into the second. For instance, the sphere is simply connected, while the torus is not simply connected.



As it turns out, the space of rotations isn't simply connected – unless we allow ourselves to represent each rotation twice, in which case we get the quaternions. To see this, consider the cycle given by starting with a 180-degree rotation around the vertical axis, and then continuously turning the axis until it points downward.



Since a 180-degree clockwise rotation is the same as a 180-degree counterclockwise rotation, our path starts and ends at the same rotation. If the space of rotations were simply connected, we would be able to smoothly adjust and contract this loop until we get a single point. However, no matter how we adjust this path of rotations, our axis of rotation must at some point be horizontal. Therefore, we cannot turn this path into a single point, and so our space is not simply connected.

The unit quaternions get around this by representing each rotation in two ways. As a result, we need to rotate our axis by a full 360° in the space of quaternions to get back to the same point (in the space of unit quaternions) we started with. Additionally, we can easily map the unit quaternions to the surface of a four-dimensional sphere and back (since the unit quaternions satisfy $q_w^2 + q_x^2 + q_y^2 + q_z^2 = 1$). Since the four-dimensional sphere, like the three-dimensional sphere[6] is simply connected, the unit quaternions themselves are simply connected, which is nice.

---

[6] Unlike the two-dimensional sphere (the edge of a circle), which essentially doesn't have enough dimensions to transform paths.

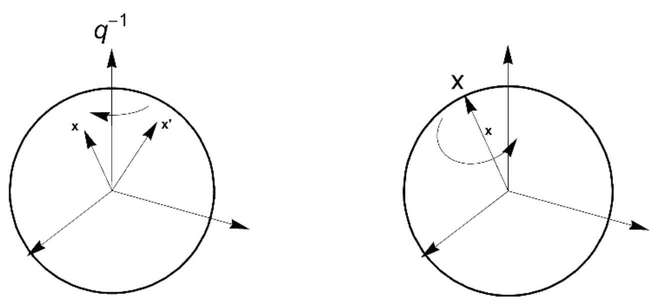
5. Working with Quaternions: 9 Recipes and Tricks You Might Not Have Heard About

Rotating points: We've talked about how the unit quaternion $\left(\cos\frac{\theta}{2}, n_x \sin\frac{\theta}{2}, n_y \sin\frac{\theta}{2}, n_z \sin\frac{\theta}{2}\right)^T$ represents a rotation, and how to find the result of performing two rotations in sequence, but we haven't actually talked about how to rotate a point around an axis.

Let's say we have a point $(x, y, z)^T$ and a quaternion $\boldsymbol{q} = \left(\cos\frac{\theta}{2}, n_x \sin\frac{\theta}{2}, n_y \sin\frac{\theta}{2}, n_z \sin\frac{\theta}{2}\right)^T$ representing the angle and axis around which we want to rotate. We can actually think of the point $\boldsymbol{x} = (x, y, z)^T$ not as a point, but as a quaternion $(0, x, y, z)$, which represents a 180° counterclockwise rotation around the axis $(x, y, z)^T$.

Now, consider the following sequence of rotations:

- Rotate by -θ around $\left(n_x, n_y, n_z\right)^T$.
- Rotate by 180° around **x**.
- Rotate by θ around $\left(n_x, n_y, n_z\right)^T$.



We'd like to figure out what rotation this corresponds to. If we apply this rotation to our desired result – the rotation of $\boldsymbol{x}$ by **q**, which we'll call **x'** – we wind up transforming **x'** to **x**, performing a rotation which leaves $\boldsymbol{x}$ in place, and finally transforming **x** back to **x'**.

On the other hand, if we started with a vector perpendicular to **x'** (let's call it **w'**), we'd transform **w'** to a vector **w** perpendicular to $\boldsymbol{x}$, negate it, and rotate it back, ultimately, to **-w**. Therefore, our sequence of rotations is the same as a 180° rotation around **x'**.

That is, if we express points in $\mathbb{R}^3$ as corresponding 180° unit quaternions, then the quaternion

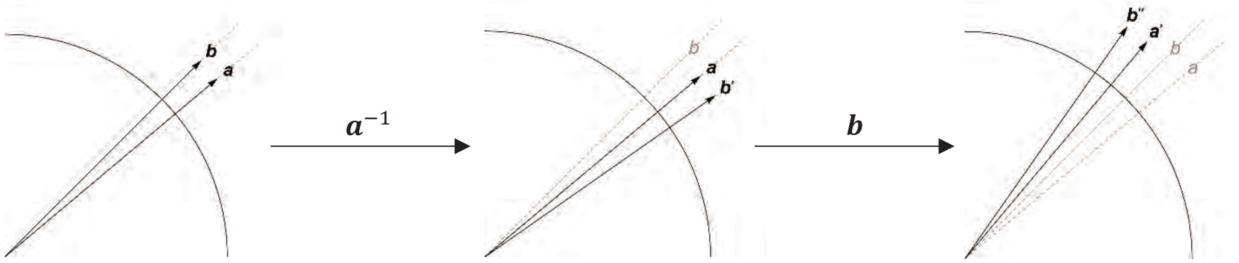
$$\boldsymbol{x}' = \boldsymbol{q}\boldsymbol{x}\boldsymbol{q}^{-1}$$

is the result of rotating the point **x** by **q**.

Rotating between two points: Suppose we have two points **a** and **b**, and we want to find some rotation which rotates the point **a** to the point **b**. Interpreting points as 180° rotations as before, consider the quaternion

$$\boldsymbol{b}\boldsymbol{a}^{-1}.$$

If we apply this rotation to (the point) **a**, we actually wind up on the far side of **b** – twice as far as we meant to go!



Intuitively (we can formalize this without much difficulty), the solution is to rotate half as far: the quaternion

$$\sqrt{\boldsymbol{b}\boldsymbol{a}^{-1}}$$

then gives a smooth and direct rotation from **a** to **b**.

We can think of the square root in much of the same way we thought of powers of quaternions at the start: the square root of an axis-angle unit quaternion corresponds to dividing its angle by 2, which we can compute either by expressing the quaternion in axis-angle form or by using the half-angle formulas for sin and cos. (We also have a sign problem from the square root; in this case, we always want to choose the quaternion with a nonnegative w component.)

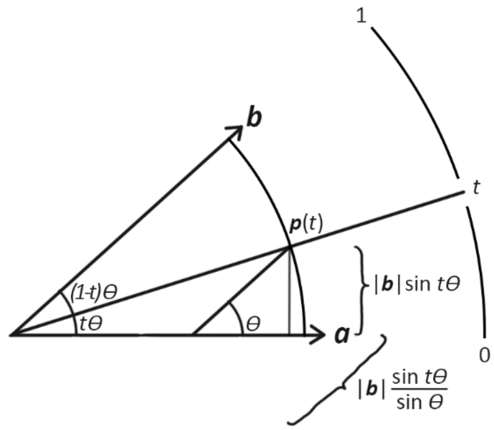
Rotations are orientations. If we have some object, we can fix some initial orientation for the object, and then describe its orientation by writing a rotation which transforms the object's initial orientation into its current orientation. Conversely, if we've fixed some initial orientation for the object, we can uniquely describe a rotation by giving the orientation of the object at the end of the rotation.

Interpolating between quaternions: Quaternions can be directly embedded in four dimensions as the set of points on the unit four-dimensional sphere. As a result, we can measure the distance between quaternions by the distance in four-space, and we can interpolate between quaternions by interpolating between points on the sphere.

In two dimensions, we can interpolate between points **a** and **b** on the unit circle (using a parameter t that ranges from 0 to 1) using the formula

$$\boldsymbol{p}(t) = \frac{\sin\big((1-t)\theta\big)}{\sin(\theta)}\boldsymbol{a} + \frac{\sin(t\theta)}{\sin(\theta)}\boldsymbol{b}$$

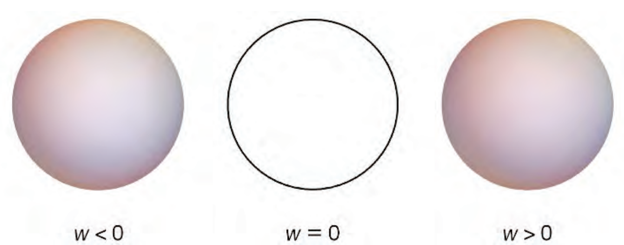
where θ is the angle between **a** and **b**. It's easiest to see this geometrically – for the component of ***b*** in the above expression, for instance:



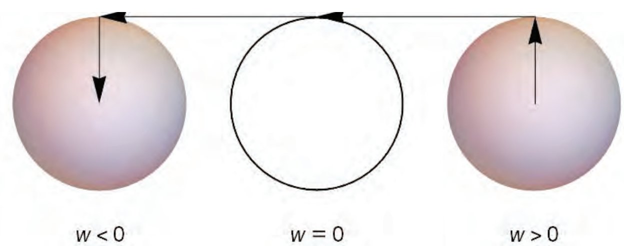
Since we can always think of a plane containing **0, a,** and **b**, the same formula also works in higher dimensions. (This is more commonly known as the *slerp* formula for spherically interpolating between two points.)

We have to be careful, though! Although this formula always gives the shortest way to interpolate between two unit quaternions, since **b** and **-b** represent the same rotation, we might have parity problems using this formula to interpolate between two rotations if we're not careful. In particular, the double-covering property of the unit quaternions combined with the above formula means that we have two ways to rotate between two rotations – and if we don't check in advance, we can wind up traversing the longer of the two ways. The solution when using quaternions to represent rotations is to choose whether to interpolate between **a** and **b** or between **a** and **-b**, depending on which of the two pairs are closer together.

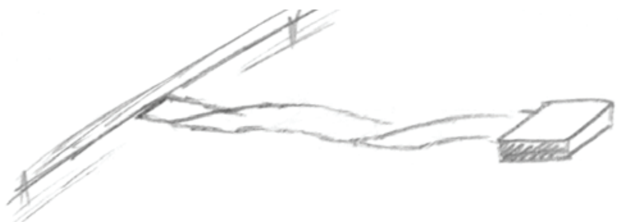
Visualizing quaternions on the sphere in four dimensions: We can lay out the surface of a four-dimensional sphere by separating it into a solid unit ball, a hollow spherical 'equator', and a second solid unit ball. The point $(w, x, y, z)^T$ on the four-dimensional sphere maps to $(x, y, z)^T$ in the left ball if $w < 0$, $(x, y, z)^T$ on the sphere if $w = 0$, and $(x, y, z)^T$ on the right ball if $w > 0$.



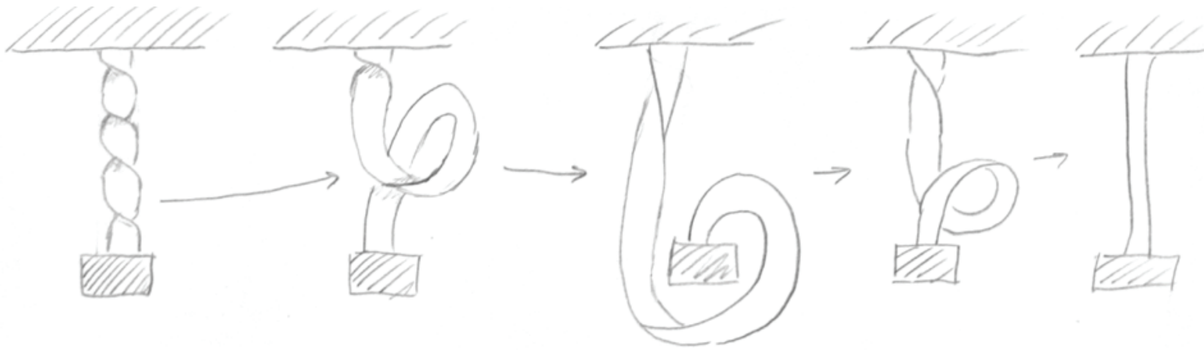
We can also draw paths between quaternions using this approach. For instance, if we were to perform a full 360-degree rotation around $(0, 0, 1)^T$ in the space of quaternions, we'd start at the center on the right (at $(1, 0, 0, 0)^T$), go upwards through the right ball, pass through the sphere at $(0, 0, 0, 1)^T$, then go downwards through the left ball, finishing in the center of the left ball at $(-1, 0, 0, 0)^T$.



The belt trick. Here's a nice way to show that the space of rotations isn't simply connected, while the space of unit quaternions is: Take a long strand of cloth (a belt will also do), fix one end of it to a stationary object, and give the end one full twist. The goal is now to find a way to bend and manipulate the middle of the cloth (possibly passing it over the end) *while keeping the ends stationary* so as to remove the twist in the cloth.



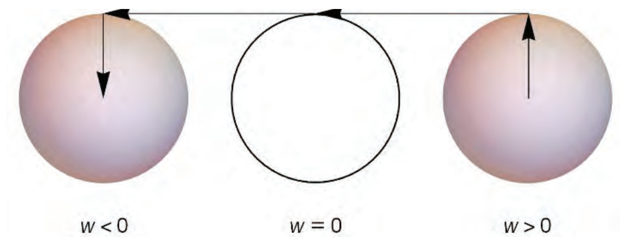
For a full twist, this is impossible – we can turn a full clockwise twist into a full counterclockwise twist, for instance, but we can't untwist the fabric. Surprisingly, though, if we start out with *two* full twists, we can untwist the fabric!



Here's the real trick: Instead of thinking about the fabric as a surface, we can think of the fabric as a series of orientations along a curve. These then form a path through the space of unit quaternions, which we can visualize!

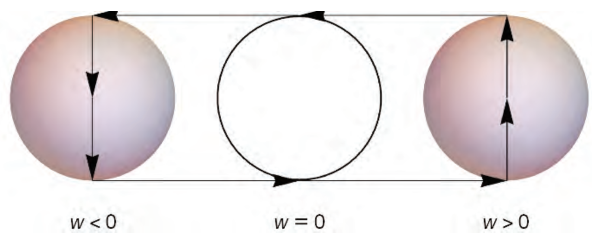
A better way, in fact, is to imagine a series of nested glass spheres around the free endpoint of the fabric, each of which contains some slice of the fabric as it leads inwards. Then we can specify the orientation and position of the fabric at any point by specifying the orientation of the corresponding glass sphere.

For a full turn, we have a series of rotations around a single axis, which gives us roughly the same path in the space of unit quaternions as before:

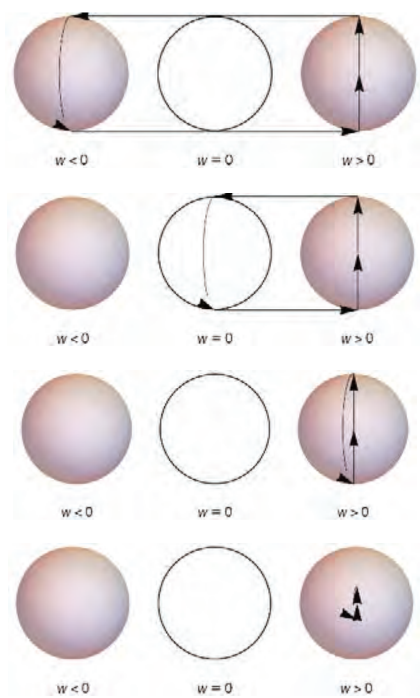


Since we start and end at different unit quaternions, though, we can't transform this path into a point while keeping the endpoints intact, so we cannot untwist the fabric.

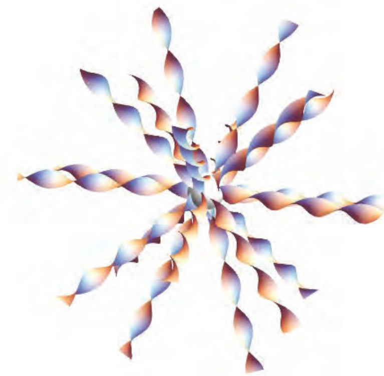
However, when we have two full twists, we have a full loop through the space of unit quaternions:



We can then untwist the fabric by transforming the path on the four-dimensional sphere to a single point:



But we have one more trick. Suppose instead of one twisted piece of fabric, we have an entire sphere of twisted pieces of fabric:



If we think of these pieces of fabric as being embedded in nested glass spheres as before, then no matter how we rotate the glass spheres, the strands of fabric will never intersect. As a result, if we start from an untwisted configuration, twist the spheres to create a double twist in one of the pieces of fabric (in fact, in all of the pieces of fabric), and then reverse the result, we'll be able to show dozens of double twists being untwisted at once without a single intersection in the entire configuration at any point.

This is quite a sight when animated; for more visualizations of this trick, see Andrew Hanson's *Belt Trick* demonstration at https://www.cs.indiana.edu/~hansona/quatvis/Belt-Trick/index.html.

Deriving quaternion composition. If we know that rotations are linear transformations, that an arbitrary rotation can be expressed as a product of rotations about the x and y axes, and that our resulting structure will require a 720-degree rotation to be returned to its initial state, we can sort of rederive the rules for composing quaternions (and in particular, the matrix at the beginning of this paper) as follows:

Let's denote our identity rotation by *1*, a 180-degree rotation about the x axis by *i*, and a 180-degree rotation about the y axis by *j*. The result of performing *j* followed by *i* is a 180-degree rotation about a third axis, which we'll denote by *k*. Since we treat a 360-degree rotation as a sort of alternate form of the identity rotation *1*, we have $i^2=j^2=k^2=-1$. We already have $ij=k$; manipulating this expression gives $jk = i$ and $ki = j$. Finally, we can see that $ji = -k$, since $ji = -(jk)(ki) = -ij = -k$, and we can similarly determine that $kj = -i$ and $ik = -j$.

Since rotations are linear, quaternion composition should be linear as well; therefore, we can split the product ***pq*** into a linear sum of ***p*****1*, ***p*****i*, ***p*****j*, and ***p*****k*. We get

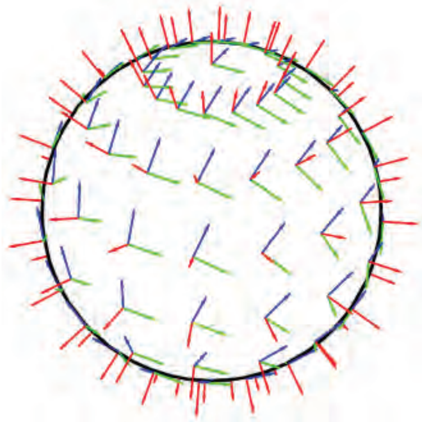
$$\begin{aligned}
(p_w + p_x i + p_y j + p_z k)1 &= p_w + p_x i + p_y j + p_z k \\
(p_w + p_x i + p_y j + p_z k)i &= -p_x + p_w i + p_z j - p_y k \\
(p_x i + p_y j + p_z k + p_w)j &= -p_y - p_z i + p_w j + p_x k \\
(p_x i + p_y j + p_z k + p_w)k &= -p_z + p_y i - p_x j + p_w k
\end{aligned}$$

so if $\boldsymbol{r} = \boldsymbol{pq}$, then

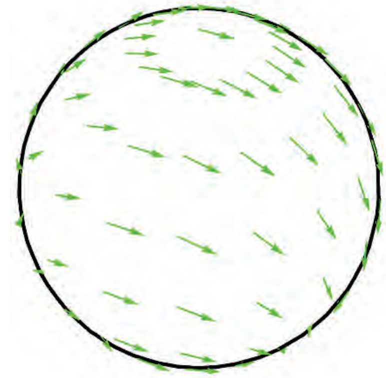
$$\begin{pmatrix} r_w \\ r_x \\ r_y \\ r_z \end{pmatrix} = \begin{pmatrix} p_w & -p_x & -p_y & -p_z \\ p_x & p_w & -p_z & p_y \\ p_y & p_z & p_w & -p_x \\ p_z & -p_y & p_x & p_w \end{pmatrix} \begin{pmatrix} q_w \\ q_x \\ q_y \\ q_z \end{pmatrix}.$$

Almost combing a hairy sphere. Although we know it's impossible to comb a hairy sphere, having some way to construct an orthonormal basis in a mostly continuous way for the points of the sphere (ideally,

from the point alone) is useful in many situations. One such method comes from [Frisvad]; basically, we can start with an orthonormal basis at the top of the sphere, and then rotate this downwards along each line of longitude, covering the entire sphere except for the south pole.



orthonormal bases



tangent vectors only

We can rotate the coordinate frame using a variety of methods; one would be to determine the rotation needed using the formula for rotating between two points above, and then to compute the result of transforming $(1,0,0)^T$ and $(0,1,0)^T$ by this rotation. If $(x, y, z)^T$ is a unit vector with $z \neq -1$, we then have that

$$\boldsymbol{b} = (1 - \frac{x^2}{1+z}, -\frac{xy}{1+z}, -x)$$

and

$$\boldsymbol{t} = (-\frac{xy}{1+z}, 1 - \frac{y^2}{1+z}, -y)$$

are perpendicular to $(x, y, z)^T$, to each other, and form the rest of a coordinate frame containing $(x, y, z)^T$ as one of its axes. When z=-1, we just output the coordinate frame $(0,0,-1)^T, (0,1,0)^T, (1,0,0)^T$.

Unfortunately, as a result of the singularity in this method, we wind up having numerical issues when z is close to -1. There are at least two approaches to solve this problem, both of which work just fine in practice. [Max] finds the optimal cutoff point for determining when we're at the south pole, which works pretty well. [Duff et al.] and [Reynolds] extend the original approach in a nice way: instead of propagating a coordinate frame downwards from the north pole, we can also propagate a coordinate frame upwards from the south pole and meet at the equator! Although we now have a discontinuity along the entire equator, we can define an appropriate coordinate frame at every point of the sphere and arrange the frames so that the frames on the southern hemisphere are just a flipped version of the frames on the northern hemisphere.

Converting to matrices and back. Even if your internal rotation representation is based on unit quaternions, rotation matrices can come in quite handy, whether you need to express a rotation in a matrix format which a computer renderer can easily read, or to constrain your rotation to satisfy some set of constraints.

Since rotations are linear transformations (in particular, the action of a quaternion $\boldsymbol{q}$ on $\mathbb{R}^3$ by $\boldsymbol{x} \rightarrow \boldsymbol{q}\boldsymbol{x}\boldsymbol{q}^{-\mathbf{1}}$ is linear), we can start out by computing $\boldsymbol{q}i\boldsymbol{q}^{-1}, \boldsymbol{q}j\boldsymbol{q}^{-1}$, and $\boldsymbol{q}k\boldsymbol{q}^{-1}$:

$$\begin{aligned}
\boldsymbol{q}i\boldsymbol{q}^{-1} &= \left(q_w^2 + q_x^2 - q_y^2 - q_z^2, 2(q_x q_y + q_w q_z), 2(q_x q_z - q_w q_y)\right)^T \\
\boldsymbol{q}j\boldsymbol{q}^{-1} &= \left(2(q_x q_y - q_w q_z), q_w^2 - q_x^2 + q_y^2 - q_z^2, 2(q_w q_x + q_y q_z)\right)^T \\
\boldsymbol{q}k\boldsymbol{q}^{-1} &= \left(2(q_w q_y + q_x q_z), 2(q_y q_z - q_w q_x), q_w^2 - q_x^2 - q_y^2 + q_z^2\right)^T
\end{aligned}$$

Since $\boldsymbol{q}$ is a unit quaternion, we then have that rotating a point by $\boldsymbol{q}$ is equivalent to multiplying by the 3x3 matrix

$$\begin{pmatrix} 1 - 2(q_y^2 + q_z^2) & 2(q_x q_y - q_w q_z) & 2(q_w q_y + q_x q_z) \\ 2(q_x q_y + q_w q_z) & 1 - 2(q_x^2 + q_z^2) & 2(q_y q_z - q_w q_x) \\ 2(q_x q_z - q_w q_y) & 2(q_w q_x + q_y q_z) & 1 - 2(q_x^2 + q_y^2) \end{pmatrix}.$$

Trying to convert back from a (numerically computed) rotation matrix to a quaternion usually results in an overdetermined system; we have seven constraints and four parameters (with a choice of sign). We can just look at the values from some subset of the matrix, for instance, or use a nonlinear least squares method to try to find optimal values for q_w, q_x, q_y, and q_z.

6. This Paper, but in Higher Dimensions

In two dimensions, rotations have one degree of freedom; there are no axes to choose, and any rotation is parameterized by its angle. In three dimensions, rotations are parameterized by an axis and an angle, for a total of three degrees of freedom. In four dimensions and higher, we cannot rely upon the axis-angle model; instead, we can decompose any rotation into a product of rotations in two-dimensional subspaces.

If you need to compute with rotations in higher dimensions, it might make the most sense to work with rotation matrices (which are the set of orthogonal matrices with determinant 1) directly. In general, in n dimensions orthogonal matrices have n^2 parameters and $n(n+1)/2$ constraints, for a total of $n(n-1)/2$ degrees of freedom.

In fact, since a single unit quaternion has three degrees of freedom and four-dimensional rotations have six degrees of freedom, it turns out we can represent a rotation in four dimensions by a pair of quaternions. In higher dimensions, we can talk about the *octonions*, an eight-dimensional algebraic system with inverses and a norm similar to the quaternions, but *without the associative property* (that is, $(ab)c$ may no longer be equal to $a(bc)$). These can then be used to represent eight-dimensional rotations, although things get complicated.

The Hairy Ball Theorem is true in all odd dimensions (where combing a sphere is well defined); however, we *can* comb the sphere in any even dimension! One easy way to do so is the following: For any vector $(x_1, x_2, \dots, x_{2n-1}, x_{2n})^T$, the vector

$$(-x_2, x_1, -x_4, x_3, \dots, -x_{2n}, x_{2n-1})^T$$

is perpendicular to $(x_1, \dots, x_{2n})^T$. In fact, in four dimensions, we can do even better, and provide a full coordinate frame: if $(w, x, y, z)^T$ is a unit four-vector, then the four vectors

$$\begin{gathered}(w, x, y, z)^T, \\ (-x, w, z, -y)^T, \\ (-y, -z, w, x)^T, \\ (-z, y, -x, w)^T\end{gathered}$$

are all of length 1 and are all orthogonal to each other. Surprisingly, these form exactly the matrix we used to represent quaternions (and is in fact a reshuffled multiplication table on 1, I, j, and k). We can also do this in two dimensions: the vectors (x, y) and (y, -x) have the same length and are perpendicular to each other. Although this cannot be done in six dimensions, while the octonions give us a way to do this in eight dimensions.

Perhaps the most surprising thing is that past eight dimensions, the space of normed division algebras just stops; the real numbers, the complex numbers, the quaternions, and the octonions are the only algebras over the reals which have a norm and allow division by nonzero numbers. This is Hurwitz' theorem, a nice proof of which can be found in Conway and Smith's *On Quaternions and Octonions*.

As it turns out, the space of rotations is never simply connected (for instance, in dimension 2, the space of rotations is diffeomorphic to a disk, which has a hole in it), but in dimensions above 2 we can always do something like we did with quaternions, double-covering the space of rotations in order to get a space which is simply connected.

7. Sources and Further Reading

Quaternions in General:

John H. Conway and Derek Smith. 2003. On Quaternions and Octonions. A K Peters, Ltd., Natick, MA, USA.

Andrew J. Hanson. 2006. Visualizing Quaternions. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Eugene Salamin. 1979. Application of Quaternions to Computation with Rotations. Stanford AI Lab, Stanford, CA, USA.

Finding Orthonormal Bases:

Tom Duff, James Burgess, Per Christensen, Christophe Hery, Andrew Kensler, Max Liani, and Ryusuke Villemin. 2017. Building an Orthonormal Basis, Revisited. *Journal of Computer Graphics Techniques (JCGT)* 6, 1 (March), 1-8.

J. R. Frisvad. 2012. Building an orthonormal basis from a 3D unit vector without normalization. *Journal of Graphics Tools 16*, 3, 151-159.

Nelson Max. 2017. Improved Accuracy when Building an Orthonormal Basis. *Journal of Computer Graphics Techniques (JCGT)*, 6, 1 (March), 9-16.

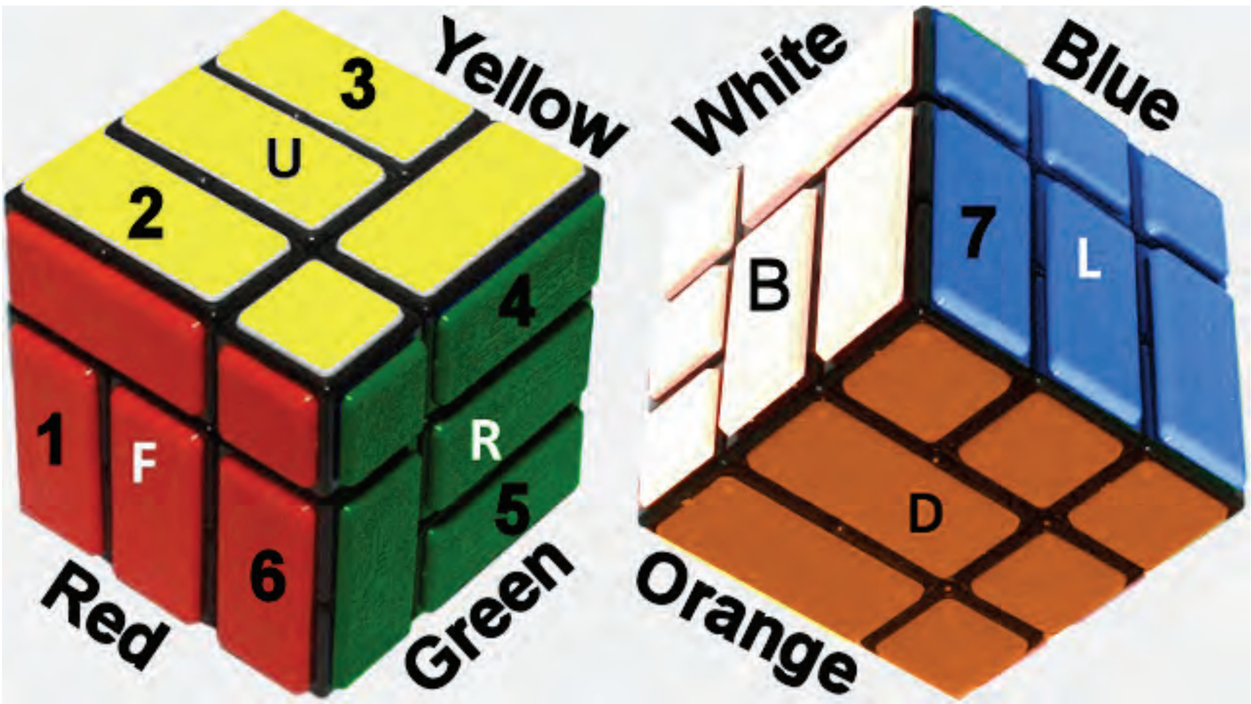
Marc B. Reynolds. 2016. Orthonormal Basis from Normal via Quaternion Similarity. URL: http://marc-b-reynolds.github.io/quaternions/2016/07/06/Orthonormal.html

The Bandaged Cube

An analysis and solution offered in tribute to
Professor Ernő Rubik & Professor David Singmaster
for the 13th Gathering for Martin Gardner.

By

Joseph Cassavaugh/Puzzles By Joe



Preface: Since not all orientations of the Bandaged Cube are equal it makes sense to add a couple of conventions to simplify the discussion of the Bandaged Cube. Above is the coloring on the cube used for this discussion. If your cube is differently colored, you'll need to make an adjustment to cover that. I will also assume that if you're interested in the Bandaged Cube that you're fairly competent with the basic Rubik's Cube's movements and notation.

Analysis and conventions:

Colors = Y, G, R opposite O, B, W respectively.

There is only one single cubelet; YGB. Looking at the cube with this cubelet in the URF corner, gives you Yellow on Up, Green on Right, Red in Front, Blue on Left, etc. This is what I will call the Home position.

There are seven other movable couplets and these will be denoted by their two long edges. GY, YR, RG (group-1), GO, RB, YW (group-2) and finally the BW piece.

The WO centers are connected – leaving just the other 4 centers to turn (YRG and B). The YRG sides are rotationally symmetric. I looked for an operator that would leave the BW piece untouched. I found the following, really nice one:

F R U F R- F2 U-

Let's call that the "Move" and label it M.

Perform M in home position and you'll notice that RG piece stays in place.

I numbered the remaining 5 edges as follows (6 and 7 being the other two).

1. **RB**
2. **YR**
3. **YW**
4. **YG**
5. **GO**
6. **RG**
7. **BW**

I think of it as just a clock-wise numbering by looking at the single-cubelet (with Yellow on top of course).

This is a nice 5-cycle 1->2->5->4->3

Doing this 5 times, of course restores the Cube.

The Move (M) can be done with Yellow, Green or Red on top (keeping the single cubelet (YGR) in the upper-front-right spot).

I knew that the 5-cycle wasn't going to be enough. At this point, I reasoned as follows. It's easy to get the centers aligned. It's easy to pop in BW (piece-7). It's easy to get one of the (Group-1) pieces in, which is then invariant under M from the right orientation (RG) (piece-6).

Using the 5-cycle, I can get one more in but that still left me with 11 unique positions to get to. There are 3 2-swaps and 4 3-cycles (2*4 + 3 = 11). Doing the 5-cycle from different directions only let me find one of the 11 (what I call the 23 (which swaps 2-3 and 4-5 (the 1 cube is now invariant by using M repeatedly until it's in place). I needed one more interesting move and I found one. I call it The "Swap" Move and I use S to note it (because it has to do with Swapping a piece in and out of the BW spot).

Here it is…The "Swap" Move:

(U2 L U- F- L- U- F) (R U) (U2 L U- F- L- U- F) (U-R-)

1->2->5->3 and 4->6 A 4-cycle and a 2-cycle.

U2 L U- F- L- U- F (accomplishes a lot but I think of it as getting the Yellow Red to go to the White-Blue spot (or 2->7). The (RU) then let's you do the move again to put 7->2. The final (U-R-) gets the cube back in its symmetrical form so the M operator can work.

With this operator (we'll call it S) and Y, G, R for the M operator with that color on top. (Y on top is still the home orientation and the numbering). We have all we need to solve the cube from any position.

23 **G2R-**
24 **RGS**
25 **YSG2Y2**
2 **GYSY2** 4-5-3
3 **GSR-Y2** 4-5-2
4 **SG2Y-** 3-5-2
5 **Y2SG2** 2-3-4

So, that's it. Here's the 5-step process I use to solve the Bandaged Cube.

1. Get the cube in the F, U & R bandaged symmetry without worrying which of the 7 bandaged pieces are located.
2. Get the BW (7) in place by using the first part of the "Swap" Move.
3. Get RG (6) in place by using M from either R and G on top.
4. Get RB (1) by doing M with Y on top (or just Y).
5. Then look at 2, 3, 4, 5.
   If one of them is in place use the corresponding formula (2, 3, 4, or 5) above.
   (You may need to do this twice (or backwards) depending on the remaining 3-cycle).
   If none of them is in place look at where the YR (2) exchanged with – either 3, 4, 5 and use that formula (23, 24, 25) from above.

Bonus Cube Fun Fact:

Q: What's special about (R2L2F2B2U2D2)(RBLFR)?

A: On each of the 6 faces 3 of the colors appear twice and 3 of the colors appear once.

I originally named any state that met these criteria as "The Most Random State" of the cube. Shortly after, I asked myself if I could create a cube that met this Most Random State criteria as well as having the property of being bilateral symmetric on all sides. I achieved that as well; but leave it as an exercise for you to enjoy if you're so inclined. I called that the Most Non-Random Random State but I never found a short algorithm to produce that. Swapping the order of the two operators above by putting the Checkerboard operator after RBLFR also results in "The Most Random State" criteria being met.

About the Author:

I am Joe Cassavaugh. I am the CEO/Designer/Engineer of Puzzles By Joe.
I graduated in 1979 with a B.S. in Mathematics from Rensselaer Polytechnic Institute.
I discovered the Rubik's Cube at the end of 1980. Although slow by today's standards, I still solve the cube in 34-ish seconds on average using a modified version of my original solving method.
I have been programming since 1981.
I have been a software engineer since 1996.
I have been a game developer since 1991.
From 2004-2009, I created the Mah Jong Quest trilogy for iWin.
I have been an Indie-Game-Dev since 2010.

I have created a fairly successful series of games for the casual PC/Mac download market called Clutter and I'm about to release the 8th game in that series. It is its own sub-genre of the Hidden Object Games genre. Although the main game mechanic is targeted at the Hidden Object crowd, there are many unique minigames which would appeal to the Recreational Mathematics crowd. In addition to those games, I have several free games from 2000-2003 that may be of interest to those of us that enjoy more logic-based puzzles. Those include Recon, a logical Battleships game (based on the puzzle game that was popularized by Games Magazine), Rack'Em, a logical Pool game and GAPWAR, a matching-edge game.

My email address is JoeCassavaugh@aol.com

If you send an email and mention the Gathering of Gardner and whether you prefer PC or Mac, I will send you a free link to the latest Clutter game.

You can also find the free logic-based windows (PC) games at:

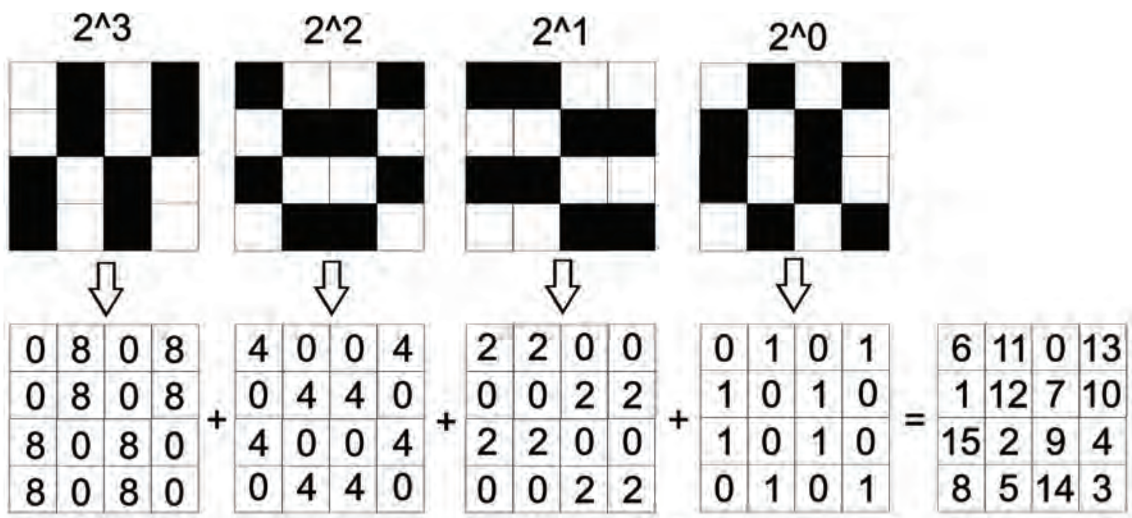
www.puzzlesbyjoe.com/other-games

11-2017 By Douglas A. Engel, Littleton, Co

Magic Squares and Space Numbers Copyright (c) 1972-2017

Jain magic square space number analysis. *Ref. link: https://en.wikipedia.org/wiki/Most-perfect_magic_square* .

The Parshvanath Jain temple in Khajuraho India has a most perfect magic square, meaning a magic square where the greatest number of possible magic sums appear. It was produced in about the year 1000. It is a 4x4 square with numbers 1 thru 16 placed in its 16 interior squares. The numbers are arranged so that all columns, main diagonals and 2x2 corner sub-squares all add up to the magic constant 34. Many other symmetrical patterns also add up to 34. It is also known as a diabolical square since it has so many ways to derive the magic constant.



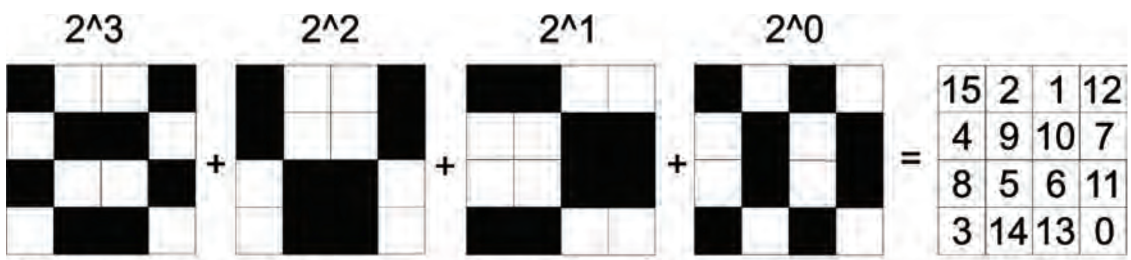
The 1000 AD Jain magic square found in a temple in India is 'most perfect' and diabolical. This space # produces 4! possible shufflings, 2^3 reverses(blk/wht exchange) (since 2^4 would include reversing all 4, equivalent to a 180 deg rotation), no rotations since the two pairs are 90 deg rotations of each other and a 180 deg rotation is the same as a reverse. This equals 24x8= 192 Jain mg'c sq's(est.). Its space # magic constant is 30.

Since it is a 4x4 square and 2^4=16, it can be seen as four two dimensional power patterns, 2^0, 2^1, 2^2, 2^3 as shown above by the black and white 4x4 cells. Think of each power pattern as having zeroes in the white squares and the power number in the black squares. Thus the 2^3 pattern has 0's and 8's. The numbers of the magic square are produced by adding the power numbers in each position together. Then the lower right Jain square numbered cell would equal 0+0+2+1=3. This representation is called a space number. Space numbers subtract 4 from the magic constant changing it from 34 to 30(numbers 0 thru 15) but it remains just as magic. Looking at the four power patterns you see immediately why it is so magical. For instance you can see that all the columns and rows have 2 black squares and two white squares and each 2x2 corner subsquare has two black squares and two white squares and similarly for the two main diagonals, and so on for the other magic patterns of the Jain magic square. In addition we can shuffle the power patterns(such as let the 2^0 and 2^2 power patterns exchange positions) however we like and the square stays most perfect, only the numbers change positions. We can also reverse the black and white of any power pattern and the magic is retained.

Rotation of individual power patterns is not allowed as the two pairs of the power patterns are already ninety degree rotations of each other. Thus with shuffling and reversing we can have 24x16=384 different number arrangements of the Jain magic square. Of course many of these will be rotations or reflections of others others, but it shows the ease with which space numbers can be used to make more magic squares. All of these different magic squares can be thought of as a single space number. The 4x4 magic square shown in Albrecht Durer's 1514 engraving 'Melancholia' can be analysed the same way and produces a very similar space number with 384 number arrangements. Other magic squares could be produced using space numbers. For instance a 12x12 magic square would need to use power patterns in base 2 and base 3.

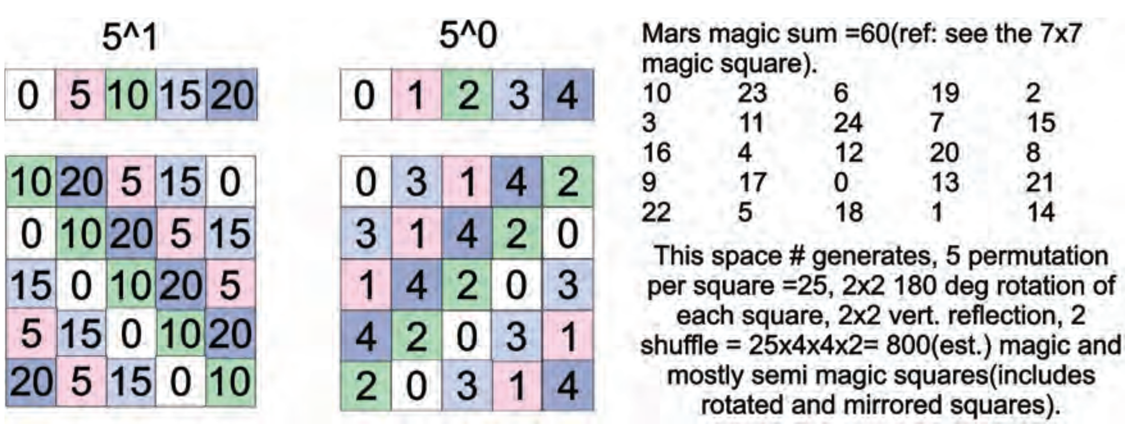
The Durer magic square as a space number

Here is an illustration of Albrecht Durer's magic square from his 1514 engraving Melancholia. It is not quite as magic as the Jain magic square but has several interesting features detailed here https://en.wikipedia.org/wiki/Magic_square#Albrecht_D.C3.BCrer.27s_magic_square
The four power patterns can be exchanged, and black and white reversed for a total of 4!x(2^4)=192 new magic squares. There may be some duplicates since reversing all four is the same as a 180 degree rotation.



Albrecht Durer,s magic square, from his engraving Melancholia as a 4x4 space number(space # magic constant = 30) produces 192(est.) Durer magic squares.

The figure below shows the Mars magic square as a base 5 space number. Since 5 is prime this is the only way it can be shown as a space number. It shows how much simpler an odd order magic square can be as opposed to an even order magic square.



Below we show the Sol magic square as a space number. Since it's prime factors are 2 and 3 it is

necessary to use both binary and trinary base numbers to break it into a space number. From this you see it has some complexity since it contains both even order and odd order properties. All even order magic squares are more complicated than odd order magic squares.

3^1 #'s x 4

0	12	24

0	24	0	24	24	0
0	0	24	24	0	24
12	12	12	12	12	12
12	12	12	12	12	12
24	24	0	0	24	0
24	0	24	0	0	24

3^0 #'s x 4

0	4	8

4	4	0	8	8	0
4	8	0	0	4	4
4	0	0	0	8	8
4	4	8	8	4	0
0	4	8	8	0	8
8	4	8	0	0	4

2^1

0	2

0	2	2	0	2	0
2	2	2	2	2	0
2	0	2	2	2	2
0	2	0	0	0	0
0	0	0	0	0	2
2	0	0	2	0	2

2^0

0	1

1	1	0	1	0	0
0	0	0	1	1	1
0	1	1	0	0	1
1	1	1	0	0	0
0	0	1	0	1	1
1	0	0	1	1	0

The 6x6 Sol magic square(see the ref. given on the binary sn 8x8). Magic constant=105. This base 2 and base 3 space # generates 1,,990,656(est.) magic and semi magic sq's. https://en.wikipedia.org/wiki/Magic_square

5	31	2	33	34	0
6	10	26	27	7	29
18	13	15	14	22	23
17	19	21	20	16	12
24	28	9	8	25	11
35	4	32	3	1	30

Next we show the Venus 7x7 magic square below. It can only be shown as a base 7 space number since 7 is prime. You can see it is very simple with a diagonal symmetry of numbers similar, to the order 5 Mars magic square.

7^1

0	7	14	21	28	35	42

21	42	14	35	7	28	0
0	21	42	14	35	7	28
28	0	21	42	14	35	7
7	28	0	21	42	14	35
35	7	28	0	21	42	14
14	35	7	28	0	21	42
42	14	35	7	28	0	21

7^0

0	1	2	3	4	5	6

0	4	1	5	2	6	3
4	1	5	2	6	3	0
1	5	2	6	3	0	4
5	2	6	3	0	4	1
2	6	3	0	4	1	5
6	3	0	4	1	5	2
3	0	4	1	5	2	6

The 1531 ed. of De Occulta Philosophia by Heinrich Cornelius Agrippa has this 7x7 magic square as planet Venus (astrology). Presented here, 11, 2017, as a base 7 space number, magic constant =168. Ref. https://en.wikipedia.org/wiki/Magic_square

21	46	15	40	9	34	3
4	22	47	16	41	10	28
29	5	23	48	17	35	11
12	30	6	24	42	18	36
37	13	31	0	25	43	19
20	38	7	32	1	26	44
45	14	39	8	33	2	27

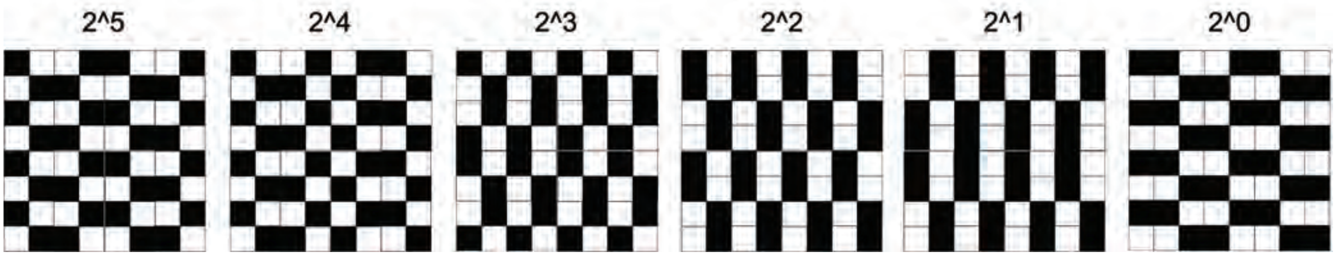
Positions for new magic and semi magic squares: 7 permutation per square =49, 2 180 rotate per square =4, 2 reflection per square =4, 2 shuffle. This gives 49x4x4x2 =1,568(est.) magic and mostly semi magic squares. For semi magic squares at least the columns and rows all add up to the magic constant.

The 8x8 magic square below is presented as a binary space number with 6 power patterns. You can see that the columns and rows all have 4 black cells and 4 white cells. For each column and row the black cells must intersect equal numbers of black and white cells and this is also true for the white cells. The same is true for the two main diagonals. This is the equal fractional intersection rule for a space number to uniquely number all the cells and be magic when combined. By considering the power pattern symmetry operations of rotations, shuffles(exchanges), reverses of black and white we get 92,160 magic squares from this space number.

Pandiagonal 8x8 magic square, sum 252, presented here as a binary space number. By Willem Barink, 4,2007.
Ref. https://en.wikipedia.org/wiki/Magic_square

61	3	12	50	45	19	28	34
4	58	53	11	20	42	37	27
51	13	2	60	35	29	18	44
10	52	59	5	26	36	43	21
63	1	14	48	47	17	30	32
6	56	55	9	22	40	39	25
49	15	0	62	33	31	16	46
8	54	57	7	24	38	41	23

Looking at the binary space number patterns you can see immediately why it is so magical. Every row, column, main diagonal, pandiagonal, 2x2 subsquare, linear group of 4, 4x4 subsquare contains equal numbers of black and white squares. The space number generates 6! possible shufflings, and 2^6 black/w color reverses which equals 46,080(est. min.) order 8 magic squares.



The same 8x8 magic square above can be analyzed with a base 8 space number. Shown below you can see it is really four 4x4 magic squares. On some permutations it probably produces many semi magic squares while the binary 8x8 space number above produces only fully magic squares. The base 8 patterns each combine 3 of the binary power patterns into one power pattern. Thus less freedom exists to move the power patterns around by shuffling and color permutation. This is partly made up for since more permutations are allowed with the base 8 pattern.

8^1

0	8	16	24	32	40	48	56

56	0	8	48	40	16	24	32
0	56	48	0	16	40	32	24
48	8	0	56	32	24	16	40
8	48	56	0	24	32	40	16
56	0	8	48	40	16	24	32
0	56	48	0	16	40	32	24
48	8	0	56	32	24	16	40
8	48	56	0	24	32	40	16

8^0

0	1	2	3	4	5	6	7

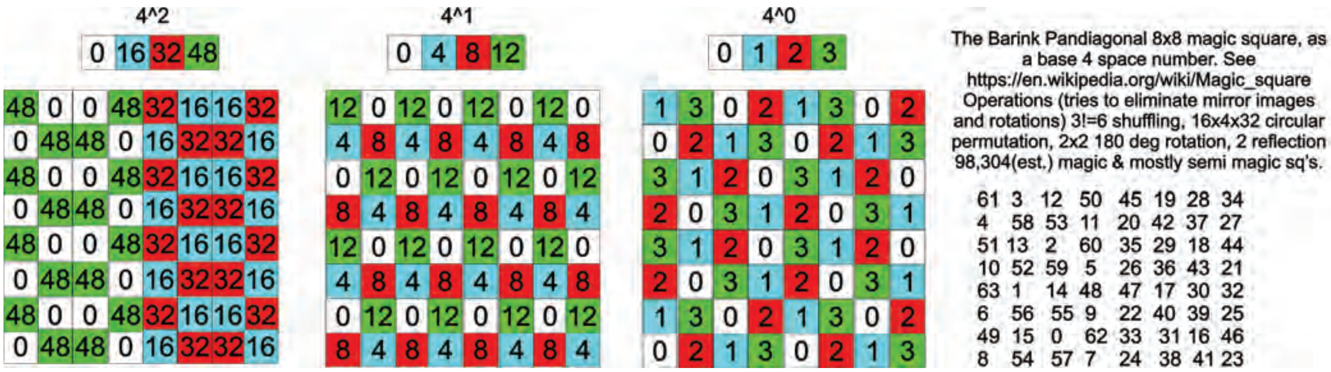
5	3	4	2	5	3	4	2
4	2	5	3	4	2	5	3
3	5	2	4	3	5	2	4
2	4	3	5	2	4	3	5
7	1	6	0	7	1	6	0
6	0	7	1	6	0	7	1
1	7	0	6	1	7	0	6
0	6	1	7	0	6	1	7

The Barink Pandiagonal 8x8 magic square, as a base 8 space number. (See the 8x8 magic square used for the binary space #) https://en.wikipedia.org/wiki/Magic_square

61	3	12	50	45	19	28	34
4	58	53	11	20	42	37	27
51	13	2	60	35	29	18	44
10	52	59	5	26	36	43	21
63	1	14	48	47	17	30	32
6	56	55	9	22	40	39	25
49	15	0	62	33	31	16	46
8	54	57	7	24	38	41	23

Color permutation =4x8x4x8, rotation =2(180 deg only), vert. reflection =2 and shuffling =2 this base 8 space # generates 4,608(est.) magic and mostly semi magic squares (does not count rotations and mirror images) .

We can also use a base 4 space number to analyze the Barink 8x8 magic square. This results in the figure below showing three 4x4 power patterns that add together to equal the 8x8 magic square. Using a different but compatible base produces a different set of magic squares where the space number symmetry operations of power shuffling, rotation, reflection and permutation are performed. It is interesting that the 4x4 pattern produces the greatest number of new magic squares.



This is but a tiny fraction of the total number of possible 8x8 magic squares which is a large number. Using space numbers this huge number can be reduced by about 10^5 since each space number representation could be taken to represent all the magic squares that can be generated from it.

Introduction to the concept of Space Numbers

Space numbers are a democracy of numbers where all numbers and dimensions have equal importance. While appearing complex Space Numbers are wonders of simplicity and symmetry.

Historical: The idea for space numbers came to me one morning around 1968, waking up at a Mining camp in the Colorado mountains, altitude over 12000 feet. I was learning to work with an IBM 360. It filled a whole room. Now you have many times the capacity of a 360 in a smart phone.

Looking at the walls there appeared different checkerboard patterns and the idea that they could combine to create numbers when overlapped. Several years later this was published as *"Number Patterns in More than One Dimension"* in the Journal or Recreational Mathematics, edited by my friend Harry L. Nelson, working at Lawrence Livermore labs in California.

Space Numbers

I have continued to work with this idea, now calling them space numbers Sn, and **Sn**() which is the set of all possible space numbers. I have devised some entertaining space number computer games. These games use rules of symmetry to recombine the space number patterns in many interesting ways to create different symmetrical and random number lists.

A Simple Description of Space Numbers and their symmetries

A space number consists of a set of power patterns each occupying an identical cellular geometry with one base position number, b^p, in each corresponding cell of each separate cellular geometry. The power patterns are added together to make a combined identical shaped cellular geometry. This is done by adding the number in identical cells of each power pattern. The cells of the combined pattern can be a linear list, like a line of square cells, or a square grid of cells or a cubic grid of cells, or a four or larger dimension grid of cells, or any symmetrical cell geometry such as the surface of a dodecahedron with each face divided into a pentagonal grid of cells. The power patterns might be rotated, reflected, permuted, and shuffled(where b^1 becomes b^3 and b^3 becomes b^1 for instance before being re-combined. These operations are only permitted, assuming we always want a complete numbering of the cells, depending on the design of the power patterns and the symmetry of the cellular geometry.

According to this description a group of cells could be randomly scattered but would still have translation(shuffle, swap, or exchange) symmetry since it could be superimposed over itself, and permutation symmetry if some path or linear set of paths connect the cells in a way that respects their power pattern symmetry. All sorts of other detailed dimensional operations within a given power pattern. are possible depending on design.

Mathematical Description
Given positive integer base b, a power pattern is a list of n numbers 0 thru n-1 filling the n cells(1 number per cell) of a symmetric geometric pattern, where you take all the ones positions in the base b number list as the b^0 power pattern, all the b^1 positions as the next power pattern up to b^(k-1), as the final power pattern. For a single base number b and a simple geometry, line, square, cube, etc., b^k=n is the total number of cells in the pattern. The power patterns consist of positive base integers and zeroes so that the integers and zeroes can be given colors where the zeroes are usually white.

A 4x4 square grid of cells contains 16 cells. Using a binary base the numbers 0 thru 15 can be gotten using four base 2 power patterns, 2^3, 2^2, 2^1, 2^0 of 4x4 cells in each pattern. The 2^3 pattern will have eight 8's and eight 0's in its 16 cells, while the 2^2 pattern will have eight 4's and eight 0's, and the 2^1 pattern will have eight 2's and eight 0's and the 2^0 pattern will have eight 1's and eight 0's. If the power patterns have symmetry then rotation, reflection, reverse of 0's and 1's(2 color permutation) are all possible. In addition exchange(shuffle, or swap) of powers is always possible since this is translation symmetry. For instance the 2^0 power pattern could exchange values with the 2^3 power pattern so that 1's and 8's in the two patterns would change places. If the power pattern numbers do not have rotation or reflection symmetry then the only allowed symmetrical operations are translation, and usually permutation of colors.

The base b could consist of a combined set of bases. For instance with a 6x6 square of 36 cells the numbers 0 thru 35 can be gotten with four power patterns by using two base 2 power patterns and two base 3 power patterns. This is done by multiplying rightmost base position number with leftmost base power numbers as required. This definition leaves open all sorts of other ways to form the number list and thus the base positions, such as complex numbers, negative numbers functions and so forth.

Linear Example
The simplest Sn examples are linear power patterns or just a list of numbers in a line. The simplest linear example employs the binary base using zeroes and ones. Write down eight binary numbers in a column. This is an element of Sn(2x2x2) shown here with decimal equivalents. A second column is shown with the decimal values of each binary pattern. Adding these decimal values for each triplet produces the column of decimal numbers. The last three columns show the power patterns separated.

Binary	Dec. equiv.	Power patterns 2^2	2^1	2^0
000 =0	000 =0	0	0	0
001 =1	001 =1	0	0	1
010 =2	020 =2	0	1	0
011 =3	021 =3	0	1	1
100 =4	400 =4	1	0	0
101 =5	401 =5	1	0	1
110 =6	420 =6	1	1	0

111 =7 421 =7 1 1 1

Notice that the binary listing(and 3 column decimal listing) forms three symmetrical columns of zeros and ones. The three columns show what the black and white colored cells look like where the 1's are generally a black or colored square and the 0's are white squares. Each of theses columns is referred to as a power pattern, Pp. The solution to the power pattern is the decimal sum of the three columns.

The left column is the 2^2 power pattern, the middle column is the 2^1 power pattern and right column is the 2^0 power pattern. Now try flipping the left column over, top and bottom. This results in a list of the numbers 0 thru 7 as follows:
4, 5, 6, 7, 0, 1 2, 3.

In fact you can rotate any of the columns and will always get the complete list of integers, just in a different or permuted order. You can also exchange columns(also called shuffle or swap). Exchange the left and right columns to get this solution. 0, 4, 2, 6, 1, 5, 3, 7. Exchange is the most powerful operation possible because it always works to produce a full list. Rotation may not work when the geometry of the numbers matches another power pattern after the rotation, causing duplicate numbers to appear. For binary patterns a simple symmetry operation is reversal of zero's and non zero's in a pattern thus the 1's become zero's and the zero's become 1's. Another symmetry operation is circular permutation. Another symmetry operation is mirror reflection.

Operations:

Shuffle Symmetry Ss (also called swap or exchange)

You can exchange(also called shuffle or swap) any two power patterns(example a^0 with b^1 become a^1, b^0) If the cells are numbered sequentially 0 thru (n^x)-1 then all possible shuffles will produce the same sequential set of numbers. Exchange is the least restrictive and therefore the most universal and powerful symmetry. The number of cells must equal n^x and n and x must be integers.

Rotation Symmetry Rs

You can rotate a power pattern about an axis of rotational symmetry if the separate power patterns have a geometric symmetry and a group symmetry of intersection. **The group symmetry/intersection rule means that each pair of power patterns intersect each other by the same fractional amount's after the symmetry operation as before the symmetry operation for a given color(base number) of their cells.** For instance a linear list of 16 binary numbered cells can can have any of its power patterns flipped end for end so that the 0's will intersect eight 0's and eight ones of each of the other power patterns and that is true for all the other intersecting pairs of power patterns.

Permutation Symmetry Ps

You can permute colors(numbers) in an n dimensional power pattern if this meets the group intersection rule above. For instance for a base 3 pattern you might have white=0 blue=1 red=2 for the 3^0 Pp, so that a permutation sends white to red blue to white and red to blue. Other kinds of group permutations may be possible depending on symmetry of the power pattern(s) and symmetry of the overall geometry of the cells. Pairs of columns or rows could be exchanged, etc.

Mirror Symmetry Ms

Mirror reflection across a line for planar patterns and plane for 3D patterns is also possible if this meets the group intersection rule stated above. In the case of symmetrical binary patterns you can just reverse zeros and non zeros (exchange places white or 0 with color or power)

Dimensional symmetry Ds with Mixed base Sn

Ds has to do with the ability to mix different bases to create symmetrical space numbers in a given

dimension. A space number with at least a rotation symmetry must have Ds. Ds is the principle that allows us to mix different base numbers to produce space numbers for any number that can be factored.

Intersection rule: *For binary power patterns any pair of power patterns must overlap to produce an equal number of the four paired numbers 00,01,10,11.* Example call the four 4x4 square power patterns A, B, C and D. Each of these is divided into a group of 8 white and 8 colored cells The white cells are zeros and the colored cells are 2^0,2^1,2^3, or 2^4. Thus A /\ B must produce the following pairs of overlapped cells(where /\ means intersection), four each of 00, 01, 10 and 11.
Then A/\B/\C must produce two groups of 000, 001, 010, 011, 100, 101, 110, 111 to satisfy the pair rule for A^B, A^C and B^C. Then A/\B/\C/\D has to produce the full 0000 thru 1111 sequential binary numbering of all 16 cells. Any exchanging or shuffling of this intersection such as D/\A/\C/\B does the same but moves the numbers around in the solution pattern. If the patterns all have the correct intersections for the operations rotation, mirror, permutation the result is a balanced set.

At this point I must end. Much more is known about space numbers. Time being available some of this mathematical and practical application knowledge will be published. Anyone wising to contribute ideas or their own discoveries can contact me by email at dengel99@aol.com.

DIFFERENCE DICE

BRIAN HOPKINS

1. Sum Dice

George Sicherman is one of the many nonprofessional mathematicians whose work was highlighted by Martin Gardner [4]. Sicherman is credited with finding alternative positive integer labels for two six-sided dice whose sums and frequencies match two standard labeled six-sided dice as shown in Table 1.

Table 1. Two pairs of labeled six-sided dice, both of whose sums are one 2, two 3s, three 4s, four 5s, five 6s, six 7s, five 8s, four 9s, three 10s, two 11s, and one 12.

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

+	1	3	4	5	6	8
1	2	4	5	6	7	9
2	3	5	6	7	8	10
2	3	5	6	7	8	10
3	4	6	7	8	9	11
3	4	6	7	8	9	11
4	5	7	8	9	10	12

An elegant approach to this problem uses generating functions or enumerators: Represent the sums of two standard labeled six-sided dice as $(x+\cdots+x^6)^2$. Sicherman's "crazy dice" arise from a different way of breaking the polynomial into factors. In particular,

$$\begin{aligned}
&(x^1+x^2+x^3+x^4+x^5+x^6)^2\\
&\quad= x^2+2x^3+3x^4+4x^5+5x^6+6x^7+5x^8+4x^9+3x^{10}+2x^{11}+x^{12}\\
&\quad= x^2\,(1+x)^2\,(1-x+x^2)^2\,(1+x+x^2)^2\\
&\quad= (x(1+x)(1+x+x^2))\,(x(1+x)(1-x+x^2)^2(1+x+x^2))\\
&\quad= (x+2x^2+2x^3+x^4)\,(x+x^3+x^4+x^5+x^6+x^8)
\end{aligned}$$

For various generalizations, see [1, 2, 3]. Here are the criteria for a polynomial $d_i(x)$ to correspond to a die with positive integer labels.

- $d_i(0)=0$ as a nonzero constant term cx^0 would indicate c faces labeled 0.
- $d_i(1)=6$ so that the sum of the coefficients matches the number of faces of the die.
- To match a count, the coefficients must be nonnegative (although factors along the way can have negative coefficients, as above).

In the example above, the flexibility in grouping factors comes from the fact that evaluating $1-x+x^2$ at $x=1$ gives 1.

2. Difference Dice

What if the sum operation is replaced by the difference? Since the dice are indistinguishable, the absolute value of the difference is a reasonable statistic to consider, as shown in in Table 2.

Table 2. Two standard labeled six-sided dice have differences six 0s, ten 1s, eight 2s, six 3s, four 4s, and two 5s.

$-$	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Are there other labelings of two six-sided dice with the same differences and frequencies? Well, increasing all labels by a constant does not change differences, so two dice labeled $\{2,3,4,5,6,7\}$ have the same difference table. Requiring the minimal label to be 1 therefore picks one representative from an infinite family of solutions.

To look for substantially different labelings, we can adapt the algebraic approach for sums:

$$\begin{aligned}
&\left(x^1+x^2+x^3+x^4+x^5+x^6\right)\left(x^{-1}+x^{-2}+x^{-3}+x^{-4}+x^{-5}+x^{-6}\right)\\
&= x^{-5}+2x^{-4}+3x^{-3}+4x^{-2}+5x^{-1}+6+5x+4x^2+3x^3+2x^4+x^5\\
&= \left(x^1+x^2+x^3+x^4+x^5+x^6\right)\left(\frac{x^1+x^2+x^3+x^4+x^5+x^6}{x^7}\right)\\
&= \frac{\left(x^1+x^2+x^3+x^4+x^5+x^6\right)^2}{x^7}\\
&= \frac{x^2\,(1+x)^2\,(1-x+x^2)^2\,(1+x+x^2)^2}{x^7}\\
&= \left(x\,(1+x)\,(1-x+x^2)^2\,(1+x+x^2)\right)\left(\frac{x\,(1+x)\,(1+x+x^2)}{x^7}\right)\\
&= \left(x^1+x^3+x^4+x^5+x^6+x^8\right)\left(x^{-3}+2x^{-4}+2x^{-5}+x^{-6}\right)
\end{aligned}$$

which leads to labeled six-sided dice shown in Table 3.

Table 3. Another pair of six-sided dice with differences six 0s, ten 1s, eight 2s, six 3s, four 4s, and two 5s.

$-$	1	3	4	5	6	8
3	2	0	1	2	3	5
4	3	1	0	1	2	4
4	3	1	0	1	2	4
5	4	2	1	0	1	3
5	4	2	1	0	1	3
6	5	3	2	1	0	2

(Note that it would not have worked to divide the other polynomial by x^7 as that would result in a mix of positive and negative exponents.) These dice are based on the same factorization as the Sicherman (sum) dice, so one might suspect that this is essentially the only possible alternative solution.

But Table 4 gives another substantially different solution.

TABLE 4. Yet another pair of six-sided dice with differences six 0s, ten 1s, eight 2s, six 3s, four 4s, and two 5s.

−	1	2	2	3	3	6
1	0	1	1	2	2	5
2	1	0	0	1	1	4
3	2	1	1	0	0	3
4	3	2	2	1	1	2
5	4	3	3	2	2	1
6	5	4	4	3	3	0

What is happening algebraically with these?

$$(x^1 + x^2 + x^3 + x^4 + x^5 + x^6)\,(x^{-1} + 2x^{-2} + 2x^{-3} + x^{-6})$$
$$= x^{-5} + x^{-4} + x^{-3} + 3x^{-2} + 5x^{-1} + 6 + 5x + 5x^2 + 5x^3 + 3x^4 + x^5$$

Unlike the symmetric situation with standard dice, where x^i and x^{-i} have the same coefficient, this pair has, for example, $3x^{-2}$ and $5x^2$. Since our operation is the absolute value of the difference, it is too restrictive to require that pairs of dice correspond to

$$x^{-5} + 2x^{-4} + 3x^{-3} + 4x^{-2} + 5x^{-1} + 6 + 5x + 4x^2 + 3x^3 + 2x^4 + x^5.$$

Instead, writing $ax^{-i} + bx^i$ as $(a + b)x^{\pm i}$, the three pairs of dice all correspond to

$$6 + 10x^{\pm 1} + 8x^{\pm 2} + 6x^{\pm 3} + 4x^{\pm 4} + 2x^{\pm 5}.$$

In fact, many others do too. An exhaustive computer search is feasible for finding all solutions matching standard six-sided dice, but a better theoretical understanding is required to handle larger cases. This leads to our closing plea.

Question: What tools allow one to find all "factorizations" of expressions such as $6 + 10x^{\pm 1} + 8x^{\pm 2} + 6x^{\pm 3} + 4x^{\pm 4} + 2x^{\pm 5}$?

References

[1] Duane Broline, Renumbering the faces of dice, *Math. Mag.* **52** (1979) 312–315.
[2] Brian Fowler and Randall Swift, Relabeling dice, *College Math. J.* **30** (1999) 204–208.
[3] Joseph Gallian and David Rusin, Cyclotomic polynomials and nonstandard dice, *Disc. Math.* **27** (1979) 245–259.
[4] Martin Gardner, Mathematical Games, *Scientific American* **238** (1978) 19–32.

Brian Hopkins, Saint Peter's University, `bhopkins@saintpeters.edu`

Gathering4Gardner13, Atlanta, April 2018

Repetitive Patterns in the Juggler Sequence

By Gabriel Kanarek for G4G13, 4/12/2018

Juggler Sequence Fast Facts!

- Integer sequence like the famous Collatz
- First publicized by Clifford Pickover in 1992
- Starting with an integer $N_0 > 0$:
    - $N_{i+1} = \lfloor N_i^{1/2} \rfloor$ if N_i is even
    - $N_{i+1} = \lfloor N_i^{3/2} \rfloor$ if N_i is odd
    - Repeat until the sequence converges to 1 (or doesn't!)

No Trivial Loops?

A trivial loop would occur when you start with a number $N_0 = x$, and after applying the Juggler function twice, end up back at x: $J(J(x)) = x$.

It doesn't matter whether x is odd or even, because we can just substitute $x \rightarrow J(x)$: the loop would oscillate between the same two values, so let's assume x is odd.

Let's call the other value $J(x) = y$, so that $x^3 = y^2 + C$, where $0 \le C < 2y + 1$. Then we know that $J(y) = x$.

We also know that y must be even, or the next number in the sequence would be larger, not smaller; therefore $y = x^2 + D$, where again $0 \le D < 2x + 1$.

Substituting back into our earlier equation, we get $x^3 = (x^2 + D)^2 + C$, and we can expand the right-hand side to get $x^3 = x^4 + 2x^2D + D^2 + C$.

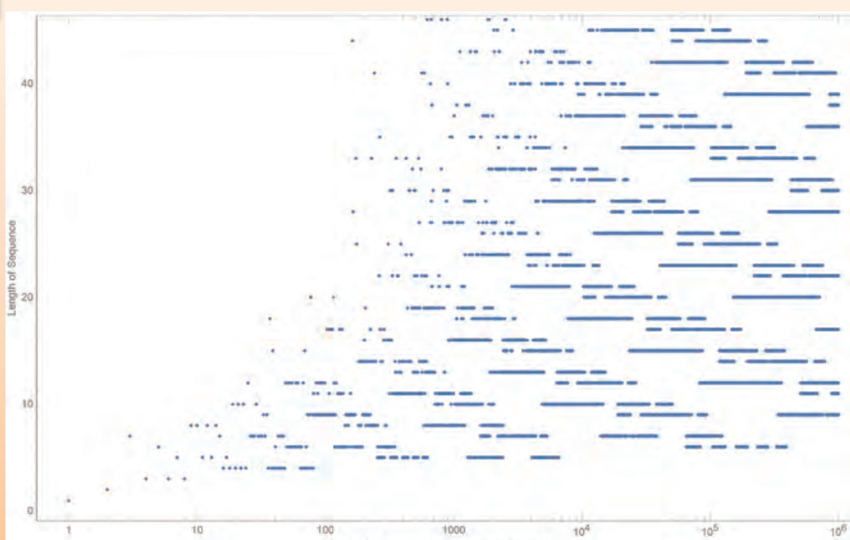
Even if $D = C = 0$, there is no integer value of $x > 1$ where this holds true, and therefore there can't be any trivial loops!

How Long Can You Juggle?

The length of Juggler sequences seems to have some patterns as well, but I haven't finished analyzing them.

They have long stretches of even numbers which all have the same length because they all get Juggled to the same number, because the Floor function puts everything between two perfect squares to the smaller perfect square.

I don't know why there are diagonal stripes. What do you think?



More Questions?

Email my dad at graykanarek@gmail.com, and he'll pass them along to me!

A Way to Derive the Spidron Formulas

Gergő Kiss
gergenium@gmail.com

Abstract: 20 years ago Dániel Erdély asked me to visualize the behavior of his invention, the (hexagonal) Spidron[1], a shape that tiles the plane, and is composed of an infinite number of increasingly small triangles. The Spidron System can be twisted up to form a periodic 3D landscape, without distorting its triangles. To visualize this, I needed to derive the equations, or formulas, that describe the transformation. This paper presents the derivation of three of the most important formulas, according to the way I followed. Later others (L Szilassi and M Hujter) used different approaches and achieved equivalent versions of the main formula presented here.

1. About the hexagonal Spidron System

A single *Spidron* is an S shaped form composed of an infinite number of increasingly small triangles (see **Figure 1**, which has three adjacent S-shapes in the middle). The two end points of the S shape are the points of convergence for the ever-shrinking triangles.

If we cut the S-shape in two, we get the so-called *Semi-Spidron* (shaded in **Figure 2**), which has only one point of convergence. There are two types of triangles in it: isosceles and equilateral ones; they follow each other in an alternating way as they converge. We can place six of these Semi-Spidrons next to each other, to get a regular hexagon. They share their points of convergence at the hexagon's center. Since the hexagon tiles the plane, the Semi-Spidron and the S-shape tile it too, creating a *Spidron System*. **Figure 1** shows three adjacent hexagons from the system.

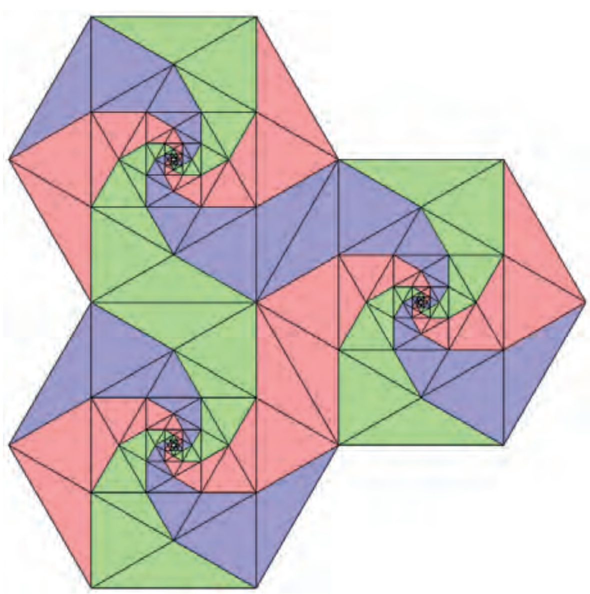


Figure 1

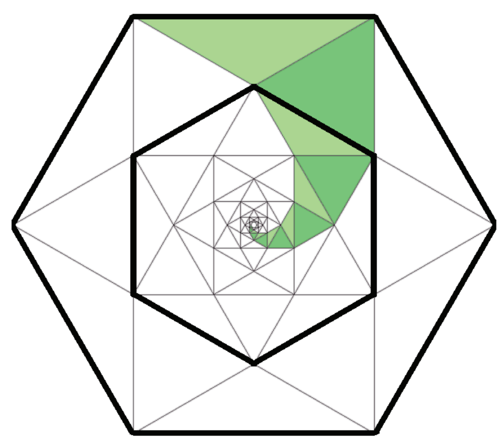


Figure 2

A whole hexagon is called a *Spidron-nest*. In the context of a Spidron-nest, a Semi-Spidron is also called a *Spidron-arm*. The cluster of triangles that form a circle (highlighted with thick black lines in **Figure 2**) is called a *Spidron-ring* (or belt).

1 Spidron™ is currently (June, 2018) a registered trade mark.

1.2 Flexibility

The hexagonal Spidron System has the ability to transform so that the shape of its triangles doesn't change. During this transformation it twists up into a periodic 3D landscape, or relief (**Figure 3**). This paper analyzes this phenomenon, by looking for the right formula, or formulas, that describe the transformation.

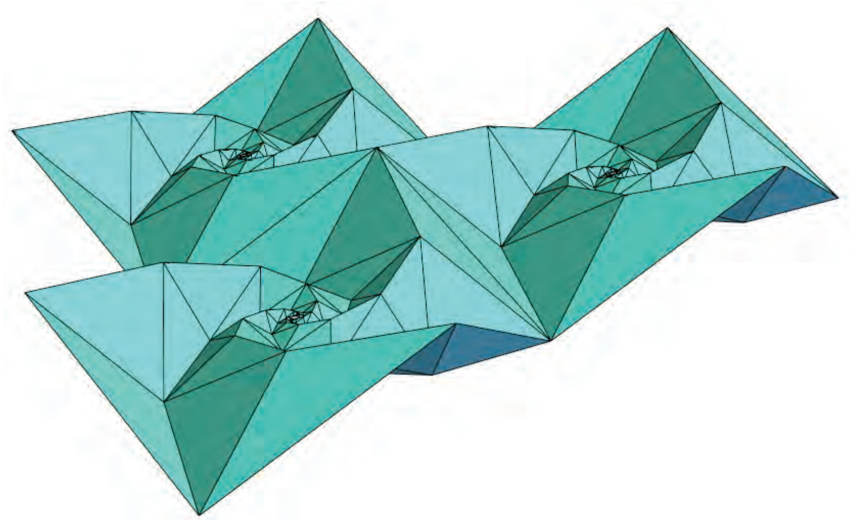


Figure 3

For further introduction to the Spidron see [1] and [2].

2. The First Formula

Let's start with the planar state of the system. Select three, mutually adjacent Spidron-nests, and concentrate on the junction point where 3 diamonds meet (highlighted on **Figure 4**).

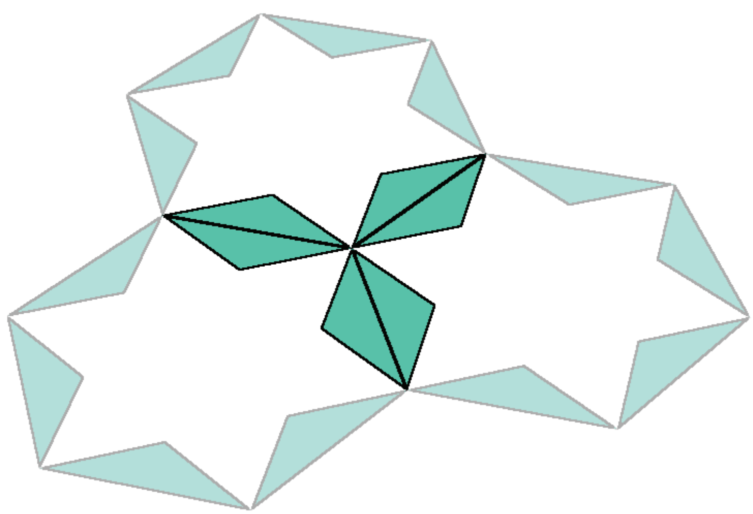


Figure 4

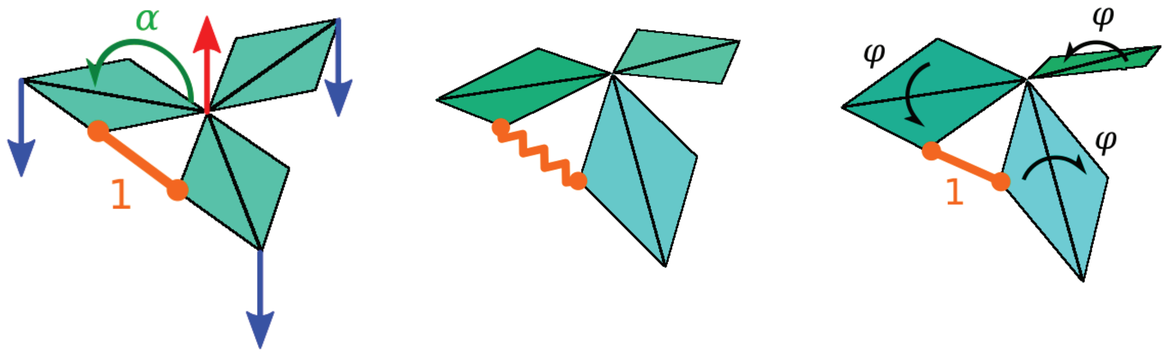
Each diamond is a pair of adjacent isosceles triangles that belong to adjacent Spidron-nests. We assume that the adjacent triangles in each pair move together, continuing to form rigid diamonds. This is a valid assumption, leading to a valid realization of the Spidron-transformation.

2.2 Leaving the plane

Let's assume that the junction point is going upwards, and the diamonds' far ends are going downwards. In terms of rotations, each diamond is getting inclined with the same α angle around axes parallel to their shorter diagonals (see **Figure 5a**).

α corresponds to the inclination angle of the hexagon's edges at the perimeter of the Spidron-nest (the lower-left one from **Figure 4**, but this applies to others in the system too[2]), relative to original plane. This is our *independent variable*. We would like to choose it freely, at least within reasonable limits.

Remember that the marked segment must retain its (say, unit) length, as there are triangles on both of its sides, which we only hid for the illustration.



a *b* *c*

Figure 5

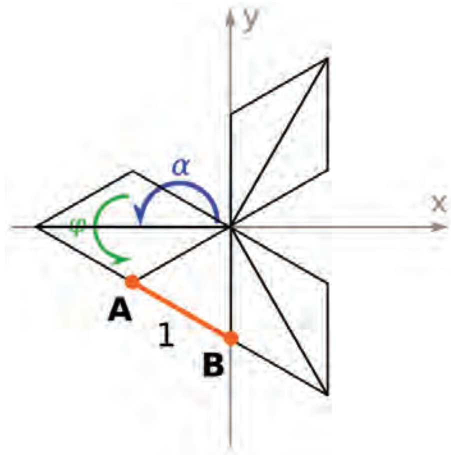
If nothing else happens, the segment cannot retain its length, it breaks (**Figure 5b**). But if we allow the diamonds to turn around their (now inclined) longer axes, with the same φ angle for each one (where the direction of the 3 rotations is symmetric, or coherent, in a circular fashion), the length of the segment can be maintained (**Figure 5c**).

This φ can, and must, be chosen well. If we manage to find a suitable value (or formula, which depends on α), the shape of the triangles in the outermost Spidron-ring will be preserved. The question is: what is this φ?

2 Sometimes I speak of a single Spidron-nest only, due the fact that the same thing happens to every Spidron-nest in the system.

2.3 Formalizing

Let’s label one end of the above segment with A, and the other with B (**Figure 6**). Their coordinates are shown. In the figure, the X and Y axes lay in the plane of the image, as indicated, and the Z axis is perpendicular to that, pointing outwards.



$$\underline{A} = \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ -\frac{1}{2} \\ 0 \end{pmatrix}$$

$$\underline{B} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}$$

Figure 6

We state that no matter what happens to the original $\underline{A}$ and $\underline{B}$ points, the distance of their transformed counterparts ($\underline{A}$’ and $\underline{B}$’) must always be 1:

$$\left|\underline{A}' - \underline{B}'\right| = 1 \qquad (1)$$

This is our main equation.

We are going to apply various rotations to the vectors around the X, Y and Z axes, by multiplying them with these known rotation matrices from the left:

$$\underline{\underline{\mathrm{Rot}_{\mathrm{X}}}}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$\underline{\underline{\mathrm{Rot}_{\mathrm{Y}}}}(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

$$\underline{\underline{\mathrm{Rot}_{\mathrm{Z}}}}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Let’s express the transformed $\underline{A}$’ (see **Figure 6**). The original $\underline{A}$ vector is first tilted by φ around the X axis, and then rotated by –α around the Y axis (α is considered to be a positive value). In the formula, these rotations follow each from right to left:

$$\underline{A}' = \underline{\underline{\mathrm{Rot}_{\mathrm{Y}}}}(-\alpha)\ \underline{\underline{\mathrm{Rot}_{\mathrm{X}}}}(\phi)\ \underline{A} \qquad (2)$$

The transformed $\underline{B}'$ is expressed very similarly, but we temporarily align its diamond with the Y axis first, by applying a rotation around Z (see **Figure 7**). This way we won't need to perform rotations around oblique axes.

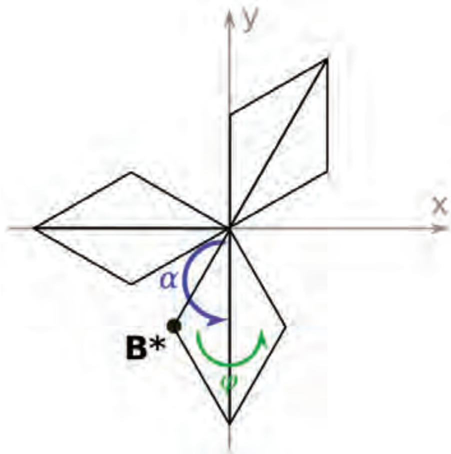


Figure 7

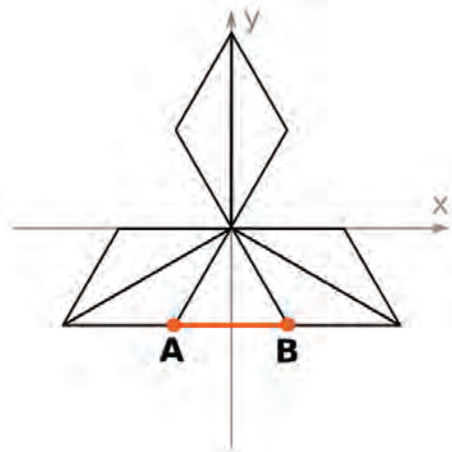


Figure 8

This correction and back-correction is valid, and doesn't interfere with the result. The Z-heading of the diamonds doesn't change when the Spidron transforms. In other words, if we project the diamonds' longer diagonals on the original plane (the plane of the image), the direction of the projections don't change during the transformation (they only get shorter). And after the previous back-correction, we too get back the original Z-heading of $\underline{B}$'s diamond.

So we first apply the aligning rotation (-30° around Z), then we tilt the aligned $\underline{B}^*$ by φ around the Y axis, and then rotate it with α around the X axis. Finally, we compensate the previous alignment applying a rotation by 30° around Z:

$$\underline{B}' = \underline{\underline{\mathrm{Rot}_{\mathrm{Z}}}}(30^\circ)\ \underline{\underline{\mathrm{Rot}_{\mathrm{X}}}}(\alpha)\ \underline{\underline{\mathrm{Rot}_{\mathrm{Y}}}}(\phi)\ \underline{\underline{\mathrm{Rot}_{\mathrm{Z}}}}(-30^\circ)\ \underline{B} \qquad (3)$$

But before going on, let's apply one more alignment, this time a permanent one. Let's rotate the whole system with 30° around Z, thus aligning $\overline{AB}$ with the X axis (**Figure 8**). This will make one of our immediate results (6) simpler (and more useful too). Because of this alignment, $\underline{A}'$ gets an additional rotation, and the back-correcting rotation of $\underline{B}'$ changes from 30° to 60°:

$$\underline{A}' = \underline{\underline{\mathrm{Rot}_{\mathrm{Z}}}}(30^\circ)\ \underline{\underline{\mathrm{Rot}_{\mathrm{Y}}}}(-\alpha)\ \underline{\underline{\mathrm{Rot}_{\mathrm{X}}}}(\phi)\ \underline{A} \qquad (4)$$

$$\underline{B}' = \underline{\underline{\mathrm{Rot}_{\mathrm{Z}}}}(60^\circ)\ \underline{\underline{\mathrm{Rot}_{\mathrm{X}}}}(\alpha)\ \underline{\underline{\mathrm{Rot}_{\mathrm{Y}}}}(\phi)\ \underline{\underline{\mathrm{Rot}_{\mathrm{Z}}}}(-30^\circ)\ \underline{B} \qquad (5)$$

These are the final formulas for $\underline{A}'$ and $\underline{B}'$.

Below is the $\overline{A'B'}$ vector (or, to be more precise, $\overline{B'A'}$ vector):

$$\underline{A}' - \underline{B}' = \begin{pmatrix} \frac{1}{2}\left(-3\cos(\alpha) + \cos(\phi)\right) \\ \frac{1}{2}\sin(\alpha)\sin(\phi) \\ -\cos(\alpha)\sin(\phi) \end{pmatrix} \qquad (6)$$

And the following is the $\overline{A'B'}$ vector's length. We want to make sure that it is 1, complying with our main equation (1):

$$\left|\underline{A}' - \underline{B}'\right| = \frac{1}{2}\sqrt{\left[-3\cos(\alpha) + \cos^2(\phi)\right]^2 + \left[4\cos^2(\alpha) + \sin^2(\alpha)\right]\sin^2(\phi)} = 1$$

If we solve this equation, φ can be expressed as a function of α like this (there are more than one solutions, but they seem to be equivalent[3] apart from differences in sign, which mean that the direction of the tilting can be reversed):

$$\phi = \arccos\left(2 - \frac{1}{\cos(\alpha)}\right) \qquad (7)$$

To achieve this formula, 20 years ago (in 1998) I used a program called *Derive*, and it needed some help from me. This year I used the *Wolfram Language*, and it solved the equation on its own.

Using this formula for φ ensures that the $\overline{AB}$ segment keeps its length, and the triangles *in the outermost Spidron-ring* don't get distorted, if we set our independent α variable to any arbitrary, reasonable value. This is the first formula.

2.4 Generalizing the angles to inner rings

You may ask: What about other, inner Spidron-rings? The above formula works for any Spidron-ring in the Spidron-nest, if we interpret it properly (we need somewhat new definitions now, which don't involve diamonds[4]): it gives the rotation angle of the ring's isosceles triangles around their longest edges (i.e. their bases), in the function of the inclination angle of these base edges relative to the original plane. These base edges are the hexagon's edges at the outer perimeter of the Spidron-ring we are currently examining.

3 My main goal was to get a working visualization, and, apart from a few numerical tests, I didn't perform thorough mathematical analysis.

4 In fact we could have operated with isosceles triangles (instead of diamonds) in the first place, concentrating on one single Spidron-nest only, but I found the idea of the junction point more intuitive because it captures more of the symmetry.

3. The Second Formula

At the end of the previous section we generalized α and φ. Let's focus more on α and its generalization. Our independent variable is called α. This is the inclination angle of the hexagon's edges at the outer perimeter of the outermost Spidron-ring (the Spidron-nest itself). There is an angle analogous to this at the outer perimeter of the *second*-outermost Spidron-ring[5]. We may call it α_2 (see **Figure 9**). What we are looking for now, is a formula that gives α_2 in the function α.

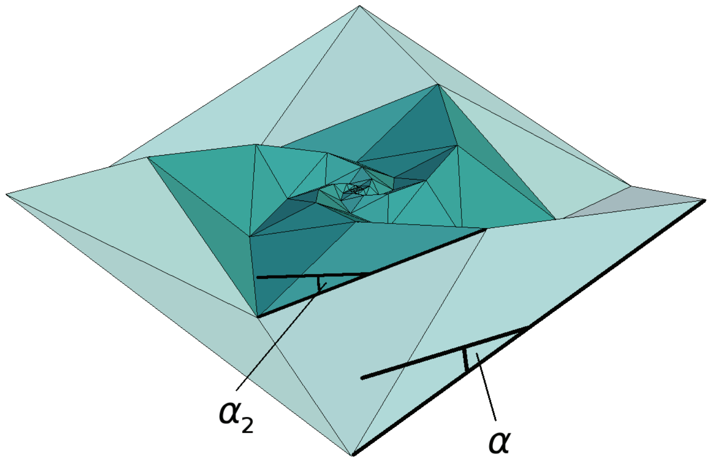


Figure 9

Since the Z coordinate of the $\overline{A'B'}$ vector (in equation (6)) represents the height-difference that B' gains over A', if we take the *arcus sinus* of this Z coordinate, we get the inclination angle of the unit-length $\overline{A'B'}$ vector relative to the XY plane, which is exactly what we are looking for. After substituting (7) in the place of φ, we get this for α_2:

$$\alpha_2 = \arcsin\left(\sqrt{-1 + 4\cos(\alpha) - 3\cos^2(\alpha)}\right) \qquad (8)$$

In fact, this formula is valid for any consecutive pair of α_{n+1} and α_n, so it provides a recursive solution to the series of α_n's. In other words, it gives the connection between the inclination angle of the hexagon's edges at the *outer* perimeter of a Spidron-ring, and the similarly defined angle at the *inner* perimeter of that ring. This is the second formula in my paper, and is generally considered to be the main Spidron Formula.

In 2004, Prof. Lajos Szilassi presented an independent derivation [3], which led to a formula equivalent to (8).

In June 2018, Mihály Hujter, PhD, described a different derivation [4], again leading to a formula which is equivalent to (8). Its English version is yet to be published at the time of writing this manuscript.

5 Which happens to be the inner perimeter of the outermost ring.

4. The Third Formula

Another important angle is the amount the second-outermost Spidron-ring turns around Z, relative to the Spidron-nest it belongs to, whose orientation we consider to be fixed.

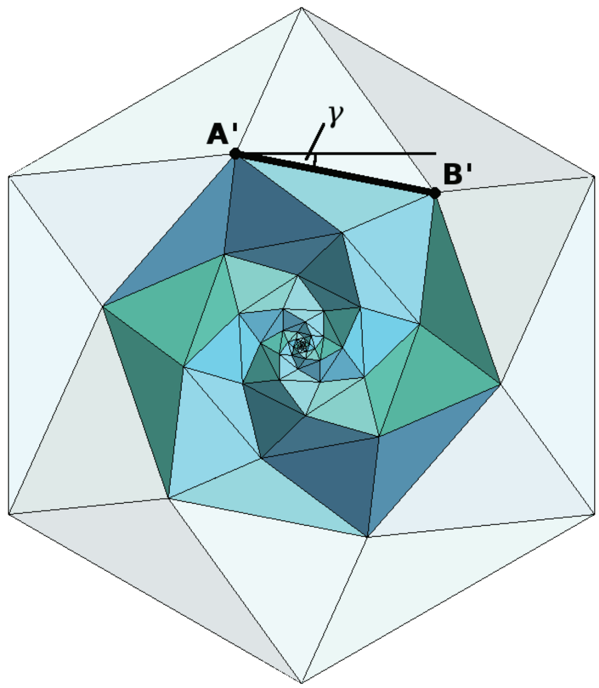


Figure 10

This is where aligning the original $\overline{AB}$ segment to the X axis becomes useful (**Figure 10**). It is now easy to read this angle (which I named γ) from the coordinates of the $\overline{A'B'}$ vector (6). We just have to take the *arcus tangent* of the ratio of its Y and X coordinates, and get:

$$\gamma = \arctan\left(\frac{\sin(\alpha)\sin(\phi)}{-3\cos(\alpha)+\cos(\phi)}\right)$$

This angle too can be generalized to inner Spidron-rings.

References

[1] Erdély, Dániel: Concept of Spidron System,
Proceedings of "Sprout-Selecting"Conference: Computer Algebra Systems and Dynamic Geometry Systems in Mathematics Teaching, Pécs, Hungary, 2004, pp. 68–77.

[2] Erdély, Dániel: Some Surprising New Properties of the Spidrons,
Bridges Conference Proceedings, 2005, pp. 179–186.

[3] Szilassi, Lajos: The right for doubting – and the necessity of doubt,
Proceedings of "Sprout-Selecting"Conference: Computer Algebra Systems and Dynamic Geometry Systems in Mathematics Teaching, Pécs, Hungary, 2004, pp. 78–96.

[4] Hujter Mihály: A csillaghatszög spidronszerű felgyűrődése mögötti matematika,
"Haladvány Kiadvány", June 16, 2018 (in Hungarian)
http://math.bme.hu/~hujter/180616.pdf

A mathematical walk around downtown Decatur, Georgia for participants at The 13th Biennial Gathering 4 Gardner Conference Friday, April 13th, 2018

Ron Lancaster
Associate Professor, Teaching Stream, Mathematics Education
Ontario Institute for Studies in Education of the University of Toronto
ron2718@nas.net

A PDF version of the math walk along with additional files can be found at https://tinyurl.com/2018G4GRL

2184: AN ABSURD (AND ADSURD) TALE

Dana Mackenzie *1165 Whitewater Cove, Santa Cruz, CA 95062, U.S.A.*
scribe@danamackenzie.com

Abstract

An old problem of De Morgan leads to the observation that the number 2184 is 3 less than a power of 3 and 13 less than a power of 13. Such a number is called "doubly absurd." Doubly absurd numbers relate in subtle ways to the Ramanujan-Nagell equation, Catalan's equation and Pillai's equation. The author conjectures that there are only seven other doubly absurd numbers, and that 2184 is the only one where both powers are cubes or higher exponents. Some partial progress is made toward proving the conjecture.

1. Introduction

Augustus De Morgan, the nineteenth century's closest analogue to Martin Gardner, once posed this puzzle: "At one point in my life, the square of my age was the same as the year." What year was he born in? It seems as if there is not enough information, until you realize that he wrote this in 1864. The only year he could have been born in (given a normal life span) was $43^2 - 43 = 1806$, so he was 43 in the year 1849.

The next birth years that could solve De Morgan's puzzle (allowing higher powers as well as squares) are 1892, 1980, 2046, 2070, 2162, 2184, 2184, ... Notice that the year 2184 appears twice! People born in the year 2184 will first be able to celebrate in 2187, when their age will be the seventh root of the year. And in case they miss this great occasion, they will get another chance ten years later: in 2197, their age will be the cube root of the year. (See [1] and [8] for popular expositions based on this idea.)

Is there some explanation for this curious fact? Is 2184 the only number that appears twice? These are the questions we will explore, and partially answer, in this paper.

First we need some terminology. If a number n can be written in the form $n = x^a - x$, we will call the number *absurd* (literally, "without the surd") because it is equal to an integer x^a minus a perfect root, or surd, of that integer. Likewise, if n

can be written as $n = x^a + x$, we will call it an *adsurd* number. If $a \geq 3$ in the above definitions, we will say that n is *strictly absurd* or *strictly adsurd*, respectively.

Furthermore, if a number n is absurd in two different ways, i.e.,

$$n = x^a - x = y^b - y, \tag{1}$$

we will call it *doubly absurd*. We can define doubly strictly absurd, doubly adsurd and doubly strictly adsurd numbers analogously. We are now prepared to say what is (probably) unique about the number 2184.

Conjecture 1. *(2184 Conjecture.) The only doubly strictly absurd number is 2184.*

Likewise, it appears that there is only one doubly strictly adsurd number.

Conjecture 2. *(130 Conjecture.) The only doubly strictly adsurd number is 130.*

We have not found a previous occurrence of the 130 Conjecture in the literature. However, the question of finding doubly absurd numbers has come up several times. The 2184 Conjecture is in fact a special case of the following more general conjecture, which is apparently due to Mike Bennett ([2], also see [9]).

Conjecture 3. *(Bennett) The eight numbers listed below are the only doubly absurd numbers.*

(i) $6 = 2^3 - 2 = 3^2 - 3$

(ii) $30 = 2^5 - 2 = 6^2 - 6$

(iii) $210 = 6^3 - 6 = 15^2 - 15$

(iv) $240 = 3^5 - 3 = 16^2 - 16$

(v) $2184 = 3^7 - 3 = 13^3 - 13$

(vi) $8190 = 2^{13} - 2 = 91^2 - 91$

(vii) $78120 = 5^7 - 5 = 280^2 - 280$

(viii) $24299970 = 30^5 - 30 = 4930^2 - 4930.$

We will call these "Bennett's eight solutions" and refer to them by number, so 2184 is Bennett's solution (v). Note that we will always write solutions to (1) with the smaller variable first, so throughout the paper we will assume $x < y$ and $a > b \geq 2$.

2. History and Heuristics

Considering that equation (1) and Conjectures 1 and 3 are not "famous," it is surprising to see what a rich history they have. In fact, there are at least three plausible routes leading to the problem: the Moret-Blanc-Mordell thread, the Ramanujan-Nagell-Skinner thread, and the Catalan-Pillai-Bennett-Mihailescu thread.

2.1. The Moret-Blanc Thread.

Is there an integer n that can be expressed both as a product of two consecutive numbers and as a product of three consecutive numbers? If we let the two numbers be $y-1$ and y, and the three numbers be $x-1, x$, and $x+1$, then we arrive at equation (1) in the special case where $a=3$ and $b=2$.

The earliest reference to this problem that we have found is [3] from 1881. Eugene Lionnet posed the above problem and Claude Seraphin Moret-Blanc, a high-school teacher in Le Havre, derived the two solutions, $n=6$ and $n=210$ (Bennett solutions (i) and (iii)).

Moret-Blanc's proof was not complete; he makes an assumption of convenience that lets him get the two stated solutions. A complete proof that 6 and 210 are the *only* solutions can be found in Mordell [4]. The proof is not elementary, as it uses the fact that a certain cubic number field has unique factorization. Thus the case $a=3, b=2$ of equation (1) has been completely solved, the only substantive case that has.

2.2. The Ramanujan Thread.

In 1913, Srinivasa Ramanujan conjectured that there are only five integer solutions to the equation

$$2^{a+2}-7=z^2. \tag{2}$$

This equation is easily reduced to a special case of (1). Clearly z must be odd, so we can write it as $z=2y-1$. Substituting this into (2) and dividing by 4, we get $2^a-2=y^2-y$, which is equation (1) with $x=2$ and $b=2$. Two of the five solutions, $(a,y)=(1,1)$ and $(2,2)$, are trivial, but the other three are not and they lead to Bennett's solutions (i), (ii) and (vi).

In 1948, Trygve Nagell proved Ramanujan's conjecture. In 1988, Chris Skinner (who at that time was a high-school student) replaced 2 with an arbitrary prime q and thus considered the equation $4q^a-4q+1=z^2$. In a remarkable paper for a teenager, or indeed a mathematician of any age, Skinner [5] showed there are only two other solutions. Combining Nagell's and Skinner's results, we conclude that Bennett's (i), (ii), (iv), (vi) and (vii) are the complete list of solutions to (1) where x is prime and $b=2$.

The main result of this paper (Theorem 1) is quite analogous to Skinner's Theorem. We will show that Bennett's (v) is the complete list of solutions to (1) where y is prime and $b = 3$. (In addition, we will have to assume a is odd).

2.3. The Catalan Thread.

In 1842, Eugene Catalan conjectured that the only consecutive numbers that are perfect powers are 8 ($= 2^3$) and 9 ($= 3^2$). That is, *Catalan's equation* $x^a - y^b = 1$ has only one positive integer solution. This was finally proved by Preda Mihailescu [6] in 2002, and is one of the landmark results in number theory so far this century.

Meanwhile, in the 1930s and 1940s, S.S. Pillai framed a generalization of Catalan's conjecture: for any constant c there are only finitely many solutions to the equation $y^b - x^a = c$. His conjecture remains open. Bennett's paper [2] proves the much more modest statement that for any two *fixed* values of x and y there are at most two solutions to Pillai's equation. Note that he allows exponents of 1, so one possibility is that the two solutions are

$$y^b - x^a = y - x = c, \tag{3}$$

which is simply another version of equation (1). In this way he arrived at his conjectured list of eight solutions to (1).

For any readers who might wish to compare this paper with Bennett's, it is somewhat tricky. Although the problems we consider are very similar, the point of view is quite different. In Bennett's paper, x and y are thought of as *parameters* and written as b and a (respectively), while a and b are thought of as *variables* and written as y and x (respectively). Also, note that his target of interest is c (in equation (3)), while ours is n (in equation (1)). For example, his Theorem 1.4 appears at first glance to say that c cannot be too small compared to y^b. In fact, though, it says that c cannot be too small compared to the *smaller* of y^b or y, which is y. Specifically, his result implies that $y < 6001c = 6001(y - x)$, or in other words $y/x > 6001/6000$.

Bennett's results are excellent in their context, but virtually orthogonal to the problems treated in this paper. Nevertheless, the Catalan thread is extremely important for our approach to equation (1). The fundamental idea is to reduce (1) to Catalan's equation. As we shall see, for $b > 3$ this reduction is not always completely successful, and it will be very convenient to call on the extensive computer work that has been done [7] to find "small" solutions of Pillai's equation.

By contrast, we have not been able to find any prior literature on Conjecture 2, doubly adsurd numbers, or the equation analogous to (1) with the minuses replaced by pluses. We will merely point out here that the only two doubly adsurd numbers less than 2 billion are $30 = 5^2 + 5 = 3^3 + 3$ and $130 = 5^3 + 5 = 2^7 + 2$. We hope that some readers will be motivated to pursue this problem further.

The main purpose of this article is to demonstrate the following theorem.

Theorem 1. *The only solution to $n = x^a - x = y^3 - y$ for which x is a positive integer, y is a prime, $a > 3$, and a is odd, is $n = 2184$.*

In fact, we will prove a generalization of Theorem 1 to all $b \leq 14$ (Theorem 2). The generalization, however, requires additional assumptions on a and y, so equation (1) is far from being completely solved even for these small exponents.

Although the proof of Theorem 1 looks technical, the main idea is quite simple. Suppose we are looking for solutions to the equation

$$x^7 - x = y^3 - y, \tag{4}$$

i.e., equation (1) with $a = 7$ and $b = 3$. We start by multiplying by x^2 to get $x^2(y^3 - y) = x^9 - x^3 = m^3 - m$, where we have defined a new variable $m = x^3$. Because m is "small" compared to m^3 and y is "small" compared to y^3, we conclude heuristically that $x^2y^3 \approx m^3$. On the other hand, if we define j to be the integer closest to m/y, so that $jy \approx m$, then $m^3 \approx j^3y^3$. Comparing these two approximate equations, we conclude that $x^2 \approx j^3$.

But what does "approximately equal" mean when the variables in question are integers? Ideally, it means the two integers differ by at most one. That is, $x^2 - j^3 = \pm 1$. But this is exactly Catalan's equation! By Mihailescu's theorem, it has the unique solution $x = 3, j = 2$. Since $jy \approx m = x^3 = 27$, it's easily seen that y must equal 13, and thus equation (4) has the unique solution $x = 3, y = 13$.

The proofs of Theorems 1 and 2 simply formalize the above argument and generalize it to other exponents a and b. The argument does not work at all if $b = 2$, because m is not sufficiently small compared to m^2. It works extremely well if $b = 3$. If $b \geq 4$ the argument works pretty well but with some complications that force us to look at Pillai's equation rather than Catalan's.

In Section 3 we will collect all the inequalities we need; this section does not involve any number theory. In Section 4 we move to the context of integers and prove Theorem 1 and its generalizations. Section 5 will offer some directions for future research.

3. Through the Eye of the Needle.

The main new tool involved in the proof of Theorem 1 is the following set of inequalities, which for the most part do not require x, y, a, and b to be integers. Only when we get to Lemma 3 will we assume that b (but not the others) is an integer; otherwise, all the variables in this section are merely assumed to be real numbers.

Lemma 1. *Given that $x^a - x = y^b - y$, $x > 1$, $y > 1$, and $a > b \geq 3$. Let $t = (a - b)/(b - 1)$ and let $m = x^{1+t}$. Finally, let $j = m/y$. Then*

$$1/x > x^t - j^b > 0. \tag{5}$$

Proof. Multiply both sides of equation (1) by x^t, to obtain

$$x^t(y^b - y) = x^{a+t} - x^{1+t} = m^b - m. \tag{6}$$

Equation (6) will be the starting point for all of our Lemmas as well as Theorem 1 and its generalizations. We start by establishing a few basic inequalities. First, because $x^t > 1$, we have $m^b - m > y^b - y$. Because $f(x) = x^b - x$ is an increasing function on $[1, \infty)$, it follows that $m > y$.

Likewise, note that $y^b - y = (x^{a/b})^b - x > (x^{a/b})^b - x^{a/b}$. By the same argument as above, $y > x^{a/b}$.

Next, we point out an important dichotomy. If $b^2 \geq a$, then

$$x^t < y^{tb/a} = y^{(ab-b^2)/(ab-a)} \leq y. \tag{7}$$

If on the other hand $b^2 < a$, then $y > x^{a/b} > x^b$. These are the possibilities referred to as Case 1 (i.e., $b^2 \geq a$ and $x^t < y$) and Case 2 (i.e., $b^2 < a$ and $x^b < y$) in Lemma 4, and I will continue to refer to them by those names throughout the paper. Notice that in either case, we can say from the first part of equation (7) that $x^t < y^{b/(b-1)}$.

Finally, note that (6) can be rewritten as follows: $y^b - y = x(m^{b-1} - 1)$. Consequently, we get the following inequality that will be used in Lemmas 3 and 4:

$$(y^b - y) < xm^{b-1} < y^b. \tag{8}$$

With the preliminaries finished, we turn to the proof of (5). We plug $m = jy$ into equation (6) to obtain

$$(x^t - j^b)y^b = (x^t - j)y = x^t y - m = x^t y - x^{1+t} = x^t(y - x) > 0.$$

This proves the right-hand side of (5). For the left-hand side, in Case 1 we have

$$(x^t - j^b)y^b < x^t y < y^2.$$

Thus

$$x^t - j^b < 1/y^{b-2} < 1/x^{b-2} \leq 1/x.$$

In Case 2 we have $x^b < y$ and $x^t < y^{b/(b-1)} \leq y^{3/2}$, since $b \geq 3$. Proceeding as above we conclude that $x^t - j^b < 1/y^{1/2} < 1/x^{b/2} < 1/x$. □

Remark. Lemma 1 formalizes the idea, stated in section 1, that $x^t \approx j^b$. Here we had to assume that $m = jy$. Lemmas 2 and 3 relax this assumption to $m \approx jy$ (in the specific sense of $\approx$ that was mentioned earlier).

Lemma 2. *Given x, y, a, b, t, and m as defined in Lemma 1, let $j' = (m-1)/y$. Then*

$$(b+1)/x > x^t - (j')^b > 0. \tag{9}$$

In fact, the right-hand side of (9) can be strengthened to:

$$x^t - (j')^b > \frac{b}{x}\left(1 - \frac{1}{y^{b-1}}\right)\left(\frac{1}{1+b/m}\right). \tag{10}$$

Proof. By Lemma 1, if $j = m/y$ then $0 < x^t - j^b < 1/x$. We note that $j' = j - 1/y$. Obviously, $0 < j^b - (j')^b$. From the Mean Value Theorem, applied to the function $f(x) = x^b$ with endpoints j and j', we have $j^b - (j')^b < bm^{b-1}/y^b < b/x$ (using (8)). Adding these inequalities, we get (9).

To improve on the lower bound, note that $(j - 1/y)^b (j + 1/y)^b < j^{2b}$, so

$$(j - 1/y)^b < \frac{j^{2b}}{(j + 1/y)^b} < \frac{j^b}{(1 + 1/m)^b} < \frac{j^b}{(1 + b/m)}.$$

We combine this with the left-hand side of (8) and do a little bit of algebra (left to the reader) to obtain inequality (10). □

Lemma 3. *Given x, y, a, b, t, and m as defined in Lemma 1, let $j'' = (m+1)/y$. Furthermore, assume that b is an integer. Then the following inequalities hold:*

(a) If $b \geq 4$ and $a \leq b^2$ (we will call this "Case 1"), then

$$\frac{1}{x}\left[b + \frac{2^b}{(\pi b)^{1/4}\sqrt{m^2 - 1}}\right] > (j'')^b - x^t > \frac{b-1}{x}. \tag{11}$$

(b) If $b \geq 4$ and $a > b^2$ (we will call this "Case 2"), then

$$(b+1)/x > (j'')^b - x^t > (b-1)/x. \tag{12}$$

(c) If $b = 3$ and $m \geq 4$, then

$$4/x > (j'')^b - x^t > 0. \tag{13}$$

Proof. We note that $j'' = j + 1/y$, where $j = m/y$. As in the proof of Lemma 2, we can apply the Mean Value Theorem, with $f(x) = x^b$ and endpoints j and j'', to show the right-hand side of inequality (11), (12) or (13).

To get the left-hand side, we note that

$$(j'')^b - j^b = \sum_{k=1}^{b} \binom{b}{k} j^{b-k} (1/y)^k.$$

(Here is where we use the assumption that b is an integer.) In many cases the leading term is the largest, so we separate it out and try to bound the rest.

$$(j'')^b - j^b = \frac{bj^{b-1}}{y} + \frac{j^{b-1}}{y} \sum_{k=2}^{b} \binom{b}{k} \left(\frac{1}{m}\right)^{k-1}. \tag{14}$$

By Schwarz's Lemma, the sum in equation (14) is bounded above by

$$\left(\sum_{k=0}^{b}\binom{b}{k}^2\right)^{1/2}\left(\sum_{k=1}^{\infty}\left(\frac{1}{m}\right)^{2k}\right)^{1/2} = \binom{2b}{b}^{1/2}\left(\frac{1}{m^2-1}\right)^{1/2}.$$

By a well-known inequality (which follows from Stirling's formula), $\binom{2b}{b} < 4^b/\sqrt{\pi b}$, and the left-hand side of inequality (11) follows.

In Case 2, $2^b \leq x^b < y < \sqrt{m^2-1}$, so inequality (12) follows. Finally, if $b = 3$, then (14) reduces to $(j'')^3 - j^3 = (3 + 3/m + 1/m^2)j^2/y$, and the expression in parentheses is less than 4 when $m \geq 4$. Part (c) follows. □

Remark. While all three of these lemmas say that, in some sense, $x^t - j^b \approx 0$, they actually provide very narrow windows or "eyes of the needle" that $x^t - j^b$ must lie in. We put this information to good use in the next section.

4. From Reals to Integers.

In this section we will assume that x, y, a and b are all integers and start investigating the consequences of the "eye of the needle" inequalities. We begin with the proof of Theorem 1, in which we are assuming that y is prime and that $b = 3$.

Proof. First note that $x = 2$ cannot be a solution because $x^a - x \equiv 2 \pmod 4$ but $y^3 - y \equiv 0 \pmod 4$. Thus we can assume $x \geq 3$. We define the integers t and m as in Lemma 1 (noticing that t is an integer because a is odd and $b - 1 = 2$). By equation (6) we see that $y|(m^3 - m)$, and therefore either $y|m$, $y|(m+1)$, or $y|(m-1)$. We will take the three cases in that order.

Suppose then that $m = jy$ for some integer $j > 1$. Then from Lemma 1,

$$1 > 1/x > x^t - j^3 > 0.$$

Because x, t, and j are integers, this case leads to a contradiction.

Next, suppose that $m = j'y + 1$ for some integer j'. By Lemma 2,

$$4/x > x^t - (j')^3 > 0.$$

Because all quantities in the equation are integers, this forces $x = 3$ and $x^t - (j')^3 = 1$. If $t > 1$, Mihailescu's theorem says that there is only one solution: $x = 3$, $t = 2$, and $j' = 2$. Then we have $a = 3b - 2 = 7$, $m = x^{1+t} = 27$ and $y = (m-1)/2 = 13$, and we recover the solution to equation (1), $3^7 - 3 = 13^3 - 13$. If $t = 1$, then we have $x = (j')^3 + 1$ and $x = 3$, which is a contradiction.

Finally, suppose that $m = j''y - 1$ for some integer $j'' > 1$. Because $x \geq 2$ and $y > x$, we have $y \geq 3$ and therefore $m \geq 5$. Thus by Lemma 3,

$$4/x > (j'')^b - x^t > 0.$$

This is only solvable in integers if $x = 3$ and $(j'')^3 - x^t = 1$. But by Mihailescu's theorem the second equation has no solution if $t > 1$. If $t = 1$ then $x = 3$ and $x = (j'')^3 - 1$, a contradiction.

Thus we conclude that $y = 13$, $x = 3$, $a = 7$ is the only integral solution to equation (1) under the conditions of Theorem 1. $\square$

It is natural to wonder whether we can use the same approach to solve equation (1), or rule out solutions, for other values of b. The answer is yes. First, we start with a strengthening of Lemma 2 that holds in the integer case. The basic idea of the proof, like everything in this section, is that the discreteness of the integers lets us turn inequalities into equalities.

Lemma 4. *Suppose that x, y, a, b, t, m and j' are defined as in Lemma 2, and are all integers. Assume $b \geq 4$ and $j' \geq 2$. Let $c = x^t - (j')^b$. Then $cx = b$.*

Proof. We already know from Lemma 2 that $cx < b+1$. Therefore it remains only to show that $cx > b - 1$. Assume, for the sake of deriving a contradiction, that $cx \leq b - 1$.

Lemma 2 also says that

$$(1 + b/m)cx > (1 - 1/y^{b-1})b \geq (1 - 1/3^{b-1})b.$$

(The latter inequality holds because $y > x$ and $x \geq 2$.) We substitute $cx \leq b - 1$ and conclude that $3^{b-1}(1 + b/m)(b - 1) > (3^{b-1} - 1)b$. This simplifies to

$$m < \frac{3^{b-1}b(b-1)}{3^{b-1} - b},$$

from which it easily follows that $m < b^2$. Because $m = x^{1+t}$, it follows that $x^t < b^2$. Then $c = x^t - (j')^b < b^2 - 2^b \leq 0$, where the first inequality uses the assumption that $j' \geq 2$ and the second uses the assumption that $b \geq 4$. This contradicts the fact proven in Lemma 3 that $c > 0$. $\square$

Theorem 2. *There are no integer solutions of $x^a - x = y^b - y$ such that $a > b$, y is prime, $(b-1)|(a-1)$, $(y-1, b-1) \leq 2$, and $4 \leq b \leq 14$.*

For example, if $b = 5$, Theorem 2 says that there are no solutions with y prime, $y \equiv 3 \pmod 4$, and $a \equiv 1 \pmod 4$.

Proof. Suppose, for the sake of deriving a contradiction, that all the conditions in Theorem 2 are true. Define t and m as in Lemma 1. Just as in Theorem 1, $y|(m^b - m)$ and m is not a multiple of y. Therefore $o(m)|(b-1)$ (where $o(m)$ is the order of $m \pmod y$). But also $o(m)|(y - 1)$ by Fermat's little theorem. Because $(b-1, y-1) \leq 2$, it follows that $o(m) = 1$ or 2. Because y is prime, the only residues

with order 1 or 2 are ± 1. Thus $m \equiv 1 \pmod{y}$ or $m \equiv -1 \pmod{y}$. Also, note that if b is even then $(b-1, y-1) = 1$, so in that case $o(m) = 1$ and $m \equiv 1$.

Accordingly we consider two cases, first where $m \equiv 1 \pmod{y}$. In this case, there is an integer $j' \geq 1$ such that $m = j'y + 1$. If $j' = 1$, then $y = m - 1 = x^{1+t} - 1$. By hypothesis, y is prime, and this is only possible if $x = 2$. By Lemma 2, $2c = cx < b + 1 \leq 15$, so $c \leq 7$. In addition, $2^t - 1 = c$, from which we conclude that c must equal 1, 3, or 7. Then $(a-b)/(b-1) = t = 1, 2$, or 4. Thus, equation (1) reduces to one of the following three possibilities: $2^{2b-1} - 2 = 3^b - 3$, $2^{3b-2} - 2 = 7^b - 7$, or $2^{5b-4} - 2 = 31^b - 31$. We leave it to the reader to show that none of these equations has an integer solution $b \geq 4$.

Thus we can assume henceforth that $j' \geq 2$, and apply Lemma 4. As in that lemma, let $c = x^t - (j')^b$. Then, by Lemma 5, $cx = b$, so we get the remarkable equation

$$x^t - (j')^{cx} = c. \tag{15}$$

If $t = 1$, then $14 \geq b \geq x = c + (j')^b \geq 1 + 2^4 = 17$, a contradiction. Thus we may assume that $t > 1$. In that case, by Mihailescu's theorem, $c = x^t - j^b > 1$. Thus $2 \leq c, x \leq 7$. It is easy to show that c and x are relatively prime, so there are only eight possibilities: $(x, c) = (2, 3), (3, 2), (2, 5), (5, 2), (2, 7), (7, 2), (3, 4)$ or $(4, 3)$.

Although we could go through the eight cases one by one, it is more interesting to give an approach with greater generality.

Suppose $x = p$ and $c = 2$, where p is an odd prime. Then $(j')^{2p} \equiv -2 \pmod{p^2}$. Thus $2^{p-1} \equiv (-2)^{p-1} \equiv (j')^{2p(p-1)} \equiv 1 \pmod{p^2}$, where the last step follows because the order of the multiplicative group $\pmod{p^2}$ is $p(p-1)$. This means that p is a Wieferich prime! Only two such primes are known: $p = 1093$ and $p = 3511$, and any further Wieferich primes are at least 4.9×10^{17}. Of course, for our proof it is sufficient to note that 3, 5, and 7 are not Wieferich primes.

Similarly, suppose $x = 2$ and $c = p$, where p is an odd prime. If t is even, the left side of equation (15) factors, and we easily get a contradiction. Thus t must be odd. Then, reducing (15) modulo p and using Fermat's little theorem, we have $(j')^2 \equiv (j')^{2p} \equiv 2^t \pmod{p}$. Thus 2^t is a quadratic residue $\pmod{p}$; hence 2 is a quadratic residue $\pmod{p}$, and by the Quadratic Reciprocity Theorem, $p \equiv 1$ or 7 (mod 8). In particular, p cannot be equal to 3 or 5.

If $x = 2$ and $c = 7$, the above congruence argument does not help. However, we have a delightful surprise: equation (15) reduces to $2^t - [(j')^7]^2 = 7$. This is just the Ramanujan-Nagell equation, discussed in section 1! In particular, we conclude that $(j')^7 = 1, 3, 5, 11$, or 181, the five solutions to the Ramanujan-Nagell equation. The possibility $j' = 1$ was ruled out earlier, and the other four possibilities are not seventh powers. Thus we have a contradiction.

The cases $x = 4, c = 3$ and $x = 3, c = 4$ lead to contradictions by similar congruence arguments (with no need to call upon advanced theorems). This completes the proof of Theorem 2 in the case where $m \equiv 1 \pmod{y}$.

Now we consider the other possibility, which is that $m \equiv -1 \pmod{y}$. In this case, there is an integer $j'' > 1$ such that $m = j''y - 1$. Recall from the first paragraph of the proof that b must be odd.

Here we can apply Lemma 3. First, if $a > b^2$, we set as usual

$$c = (j'')^b - x^t \tag{16}$$

and note that $b + 1 > cx > b - 1$. Thus $cx = b \leq 14$. By Mihailescu's theorem $c > 1$. Thus b is an odd composite number less than 14, which means that $b = 9$ and $c = x = 3$. But then $(j'')^b = 3^t + 3$, which is impossible because the right-hand side is divisible only once by 3.

Hence we can assume that $a \leq b^2$, and we note that this also means that $t \leq b$. It is easy to rule out $t = b$ because we can factor the right-hand side of (16) (details again left to the reader). Thus we can assume that $t \leq b - 1$.

At this point, the argument gets a little bit messy because we have to use the upper bound in Lemma 3(a), which is much worse than the one in Lemma 3(b). However, we also have the benefit of massive computer calculations that have been done to identify small solutions (in c) to the equation $(j'')^b - x^t = c$, Pillai's equation. Specifically, sequence A076427 in the Online Encyclopedia of Integer Sequences, and the linked table at [7], lists all of the solutions to this equation for which $c \leq 100$ and for which the two powers, $(j'')^b$ and x^t are less than 10^{18}. As it turns out, the table is sufficient to solve our problem for $b \leq 13$.

We will leave the easier cases, $b = 5$, 7, and 9, to the reader and give the proof in the two most difficult cases, $b = 11$ and 13.

Suppose that $b = 11$. Note that $x = 2$ can never be a solution to (1) when b is odd, by a congruence argument (mod 4). Thus $x \geq 3$ and $y \geq 5$ (remembering that y is a prime greater than x). We also know that $m \geq 2y - 1 \geq 9$. Applying equation (11), we conclude that $(j'')^b - x^t = c$, where $cx < 105.5$. Because $x \geq 3$, we have $c \leq 35$, which puts it within the range of table [7]. Also because $c \geq 2$, we have $x \leq 52$, so $x^t \leq 52^{10} < 1.4 \times 10^{17}$, so x^t and $(j'')^b$ are also within the range of the table. Thus equation (16) must appear among the 274 known solutions of Pillai's equation in table [7]. But none of those solutions involves an eleventh power, so we have a contradiction.

The case $b = 13$ requires a little extra work. First, we can rule out $x = 3$ by reducing equation (1) modulo 9, and we can rule out $x = 4$ by reducing equation (1) modulo 8. Hence any solution to (1) must have $x \geq 5$, $y \geq 7$ and $m \geq 13$. Now Lemma 3 says that $cx \leq 263.1$. Because $x \geq 5$, $c \leq 52$, which puts c within the range of table [7]. Also, if $x \leq 31$, then $x^t < 31^{12} < 10^{18}$, which puts x^t and $(j'')^b$ within the required range as well.

Now if $x \geq 32$, then y is a prime greater than x, so $y \geq 37$ and $m \geq 73$. We can recompute the upper bound (9) with the new value of m and we find that $cx < 44.4$. This is already a contradiction, because $c \geq 2$ and $x \geq 32$.

Thus $x \leq 31$ and so c, x, and j'' must appear among the 274 known solutions of Pillai's equation in table [7]. However, none of those solutions involve a thirteenth power, and so we arrive at our final contradiction. $\square$

While the last part of the proof of Theorem 2 is inelegant, the main point is that, when $m \equiv -1 \pmod{p}$, we were able to place an upper bound on t, depending on b. Thus the search for solutions was reduced to a finite, albeit large, calculation. Fortunately, that calculation was already done for us!

5. Ideas for Future Study.

In this final section, I will leave the reader with two unsolved problems.

1) Prove Conjecture 2. That is almost certainly too hard, but it would be interesting to see if analogues of Theorems 1 and 2 can be proven for doubly adsurd numbers.

2) Prove Theorem 1 without the assumption that a is odd, or prove Theorem 2 without the extra assumptions that $(b-1)|(a-1)$ and $(y-1, b-1) \leq 2$. As it turns out, when $b = 3$ the first two even cases are quite easy, but the arguments do not seem to generalize to even numbers $a \geq 8$.

Theorem 3. *If $a = 4$ or 6, then the equation $x^a - x = y^3 - y$ has no integer solutions for which $x > 1$ and y is a prime.*

In fact, the proof given below works just as well with the weaker hypothesis that x and y are relatively prime.

Proof. If $x^6 - x = y^3 - y$ and $x, y > 1$, note that $f(y) = y^3 - y$ is increasing, with $f(x^2) < x^6 - x$ and $f(x^2+1) > x^6 - x$. Hence $x^2 < y < x^2 + 1$, which is impossible if y is an integer.

The proof for $a = 4$ starts in the same way. Suppose that $x^4 - x = y^3 - y$, where y is prime. As above, it is easy to verify that $x^{4/3} < y < x^{4/3} + 1$. Also notice that

$$(y - x)(y^2 + xy + x^2 - 1) = y^3 - x^3 - y + x = x^4 - x^3,$$

from which we conclude $x^3|(y - x)(y^2 + xy + x^2 - 1)$. Because $x < y$ and y is prime, it follows that x is relatively prime to y, and hence x^3 is relatively prime to $(y - x)$. It follows that $x^3|(y^2 + xy + x^2 - 1)$. Now suppose, for the sake of deriving a contradiction, that $x \geq 8$, so that $x^{-1/3} \leq 1/2$. Then, because $y < x^{4/3} + 1$,

$$\frac{1}{x^3}(y^2 + xy + x^2 - 1) < x^{-1/3} + x^{-2/3} + x^{-1} + 2x^{-5/3} + x^{-2} < 1.$$

It is impossible for $(y^2 + xy + x^2 - 1)$ to be a multiple of x^3 and yet be less than x^3. So, by contradiction, we conclude that $x < 8$. But it is easy to verify that $y^3 - y = x^4 - x$ has no integer solution if $x = 2, 3, 4, 5, 6$, or 7. $\square$

References

[1] M. Parker, Why 1980 was a great year to be born... but 2184 will be even better (video). http://www.youtube.com/watch?v=99stb2mzspI (published June 29, 2015).

[2] M. Bennett, On some exponential equations of S. S. Pillai, *Canad. J. Math.* **53** (2001), 897-922.

[3] Solutions de questions proposees dans les Nouvelles annales, *Nouvelles annales de mathematiques 2e serie* **20** (1881), 425-432. (Moret-Blanc's solution on pp. 431-432.)

[4] L. Mordell, *Diophantine Equations*, Academic Press, Cambridge (UK), 1969, p. 257.

[5] C. Skinner, The Diophantine Equation $x^2 = 4q^n - 4q + 1$, *Pacific J. Math.* **139** (1989), 303-309.

[6] P. Mihailescu, Primary Cyclotomic Units and a Proof of Catalan's Conjecture, *J. Reine Angew. Math.* **572** (2004), 167-195.

[7] T. Noe, Number of solutions to Pillai's equation $a^x - b^y = n$, with $a > 0$, $b > 0$, $x > 1$, $y > 1$. On-Line Encyclopedia of Integer Sequences, http://oeis.org/A076427. (Posted Oct. 11, 2002.) Also see supportaing documentation at http://www.sspectra.com/Pillai.txt. (Accessed May 18, 2016.)

[8] G. Antonick, The Doubly Good Year. *New York Times* "Numberplay" blog, January 4, 2016. http://wordplay.blogs.nytimes.com/2016/01/04/the-doubly-good-year.

[9] W. Wu, Integer Solutions of $x + y^n = y + x^m$ for $n < m$.
http://mathoverflow.net/questions/158071 (posted February 19, 2014).

Playing nice with a crooked coin

1 A motivating problem

Two kids have found a coin and want a fair way of deciding who gets to keep it, by tossing it a finite number of times. Let p be the probability of the coin landing heads. We may assume by symmetry that $0 < p \leq \frac{1}{2}$. Suppose the coin is fair, that is, $p = \frac{1}{2}$. Clearly one toss will suffice.

Suppose the coin is crooked. We may toss the coin twice. A head followed by a tail means that the first kid gets the coin, while a tail followed by a head means that the second kid gets the coin. If the coin lands heads both times or tails both times, the process is repeated. This is clearly fair. However, if some super-being is having fun with the kids, they may be tossing heads until the end of time. Thus this is not a solution as it violates the condition that a fair decision must be reached within a finite number of tosses.

If $0 < p < \frac{1}{2}$, the task is not always possible, but there are infinitely many values of p for which workable protocols exist. Suppose we toss the coin twice. Two tails means that the first kid gets the coin. Otherwise the other kid gets it. To make this fair, we need $(1-p)^2 = \frac{1}{2}$. Hence $1 - p = \frac{1}{\sqrt{2}}$ and $p = 1 - \frac{1}{\sqrt{2}}$.

For another possible value, suppose we toss the coin three times. The first kid gets the coin if and only if it lands tails all three times. Then $(1-p)^3 = \frac{1}{2}$ and $p = 1 - \frac{1}{\sqrt[3]{2}}$. It is clear that $p = 1 - \frac{1}{\sqrt[n]{2}}$ works for any positive integer n.

2 A second example, introducing Pascal's triangle

Let's look at four kids now. In this example we give an idea of a general technique to produce solutions using Pascal's triangle. We present three possible solutions.

1. The obvious value is $p = \frac{1}{2}$, and the coin needs to be tossed only twice. Flipping the coin twice yields the possible results of HH, HT, TH, TT of equal probability. Assigning each result to a person provides a fair game.

2. Suppose the coin is not fair. Let $q = 1 - p$. Tossing it six times, we have Let $q = 1 - p$. Suppose we flip the coin six times. Then,

$$1 = (p+q)^6$$

so after expanding we have,

$$1 = p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6\,. \qquad (1)$$

Next we collect terms divisible by 3 together

$$1 = 3(2p^5q + 5p^4q^2 + 6p^3q^3 + 5p^2q^4 + 2p^5q) + p^6 + 2p^3q^3 + q^6\,.$$

For each of the three copies of the outcomes given by $2p^5q + 5p^4q^2 + 6p^3q^3 + 5p^2q^4 + 2p^5q$ we can assign to one person. This way we guarantee each of them will have the same probability of winning. Therefore it suffices to make sure the final person has the same probability as the others, ie we want to find find p and q such that

$$\frac{1}{4} = p^6 + 2p^3q^3 + q^6$$

or

$$\frac{1}{4} = (p^3 + q^3)^2 .$$

Defining $0 < r < \frac{1}{2}$ as $p = \frac{1}{2} - r$ yields

$$\frac{1}{4} = \left(\left(\frac{1}{2} - r \right)^3 + \left(\frac{1}{2} + r \right)^3 \right)^2$$

$$\frac{1}{4} = \left(\frac{1}{4} + 3r^2 \right)^2 ,$$

which lets us solve $r = \frac{1}{\sqrt{12}}$.

3. Similarly tossing a different crooked coins nine times, $1 = ((\frac{1}{2} + r) + (\frac{1}{2} - r))^9$ is the sum of

$$\left(\frac{1}{2} + r \right)^9 + 3 \left(\frac{1}{2} + r \right)^6 \left(\frac{1}{2} - r \right)^3 + 3 \left(\frac{1}{2} + r \right)^3 \left(\frac{1}{2} - r \right)^6 + \left(\frac{1}{2} - r \right)^9$$

and other terms whose coefficients are all multiples of 3. So we set

$$\frac{1}{4} = \left(\left(\frac{1}{2} + r \right)^3 + \left(\frac{1}{2} - r \right)^3 \right)^3 = \left(\frac{1}{4} + 3r^2 \right)^3 .$$

It follows that $\frac{1}{4} + 3r^2 = \frac{1}{\sqrt[3]{4}}$ so that $r = \sqrt{\frac{4 - \sqrt[3]{4}}{12 \sqrt[3]{4}}}$.

In summary, three possible values are $p = \frac{1}{2}$, $p = \frac{1}{2} - \frac{1}{\sqrt{12}}$ and $p = \frac{1}{2} - \sqrt{\frac{4 - \sqrt[3]{4}}{12 \sqrt[3]{4}}}$.

Remark.
Where did the inspiration to check 6 flips and 9 flips come from? The solution above is based on Pascal's Triangle, involving a subtraction of the 2nd row from the 6th row, and a subtraction of the 3rd row from the 9th row. In particular, for the second solution ($p = \frac{1}{2} - \frac{1}{\sqrt{12}}$) the numbers in the 6th row not circled are already divisible by 3, and subtracting the circled numbers in the 2nd row from the 3rd yields numbers which are divisible by 3.

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

In this sense, the 6th row "minus" the 2nd row yields only numbers divisible by 3. Because of this we are able to to rewrite (1) as a perfect square plus several terms with coefficients divisible by 3. This is our central strategy. Looking at more rows of Pascal's triangle we see this approach works again to get the probability $p = \frac{1}{2} - \sqrt{\frac{4-\sqrt[3]{4}}{12\sqrt[3]{4}}}$.

$$
\begin{array}{ccccccccccccccccccc}
 & & & & & & & & & 1 & & & & & & & & & \\
 & & & & & & & & 1 & & 1 & & & & & & & & \\
 & & & & & & & 1 & & 2 & & 1 & & & & & & & \\
 & & & & & & 1 & & 3 & & 3 & & 1 & & & & & & \\
 & & & & & 1 & & 4 & & 6 & & 4 & & 1 & & & & & \\
 & & & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & & & \\
 & & & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & & & \\
 & & 1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 & & \\
 & 1 & & 8 & & 28 & & 56 & & 70 & & 56 & & 28 & & 8 & & 1 & \\
1 & & 9 & & 36 & & 84 & & 126 & & 126 & & 84 & & 36 & & 9 & & 1
\end{array}
$$

3 A third example, finding infinite solutions

What about an arbitrary number of kids? Call it n. If $n = 3$ we need to take the difference of two rows of Pascal's triangle to obtain only multiples of 2. This should be easier to do!

Let's consider the second row of Pascal's triangle,

$$1 = \left(\frac{1}{2} - r\right)^2 + 2\left(\frac{1}{2} - r\right)\left(\frac{1}{2} + r\right) + \left(\frac{1}{2} + r\right)^2,$$

so that removing the even terms and letting the leftover equal to 1/3 gives,

$$\frac{1}{3} = \left(\frac{1}{2} - r\right)^2 + \left(\frac{1}{2} + r\right)^2$$

so that,

$$-\frac{1}{12} = r^2$$

which has no real solutions. While we weren't lucky this time, this approach **does** work for flipping the coin more than twice!

Subtracting the fourth row from the second provides leaves of 2, which produces the probability of

$$p = \frac{1}{2} - \sqrt{\frac{1}{2\sqrt{3}} - \frac{1}{4}}.$$

Subtracting the sixth row from the third also leaves multiples of 2, this time producing the probability of

$$p = \frac{1}{2} - \sqrt{\frac{1}{2\sqrt[3]{3}} - \frac{1}{4}}.$$

It seems reasonable to conjecture that for 3 people, and with $k \geq 2$, if you flip the coin $2k$ times then setting the probability

$$p = \frac{1}{2} - \sqrt{\frac{1}{2\sqrt[k]{3}} - \frac{1}{4}}$$

will provide a setup for a fair game. It is worth noting that as $k \to \infty$ then $p \to 0$.

One can show this conjecture holds true as

$$\binom{2a}{2b} \equiv \binom{a}{b} \pmod 2,$$

and

$$\binom{2a}{2b+1} \equiv 0 \pmod 2.$$

Similarly, for $n = 5$ people, one can apply a similar strategy. For any $k \geq 2$ we flip the coin $4k$ times and use a probability of,

$$p = \frac{1}{2} - \sqrt{\frac{1}{2\sqrt[2k]{5}} - \frac{1}{4}}.$$

This time the tricky part is proving the following lemma,

Lemma. *For positive integers a and b,*

(a)

$$\binom{4a}{2b+1} \equiv 0 \pmod 4 \quad \text{for } 0 \leq b \leq 2a-1,$$

(b)

$$\binom{4a}{2b} \equiv \binom{2a}{b} \pmod 4 \quad \text{for } 0 \leq b \leq 2a.$$

Proof. First we make repeated use of the recursive formula. Then for $0 \leq c \leq 4a$,

$$\begin{aligned}
\binom{4a}{c} &= \binom{4a-1}{c} + \binom{4a-1}{c-1} \\
&= \binom{4a-2}{c} + 2\binom{4a-2}{c-1} + \binom{4a-2}{c} \\
&= \binom{4a-3}{c} + 3\binom{4a-3}{c-1} + 3\binom{4a-3}{c-2} + \binom{4a-3}{c-3} \\
&= \binom{4a-4}{c} + 4\binom{4a-4}{c-1} + 6\binom{4a-4}{c-2} + 4\binom{4a-4}{c-3} + \binom{4a-4}{c-4}.
\end{aligned}$$

So we have shown
$$\binom{4a}{c} \equiv \binom{4(a-1)}{c} + 2\binom{4(a-1)}{c-2} + \binom{4(a-1)}{c-4} \pmod 4\,. \tag{2}$$
We will prove (a) first. Notice it is trivially true for $b = 0$ (equivalently for $b = 2a - 1$), and for $b = 1$ (equivalently for $b = 2a - 2$) we have

$$\binom{4a}{3} = \frac{(4a)(4a-1)(4a-2)}{3 \cdot 2} \equiv 0 \pmod 4\,.$$

We prove the rest by induction on a. The base case is easy to see. We may assume (a) is true for $a - 1$. By (2) with $c = 2b + 1$ we have,

$$\binom{4a}{2b+1} \equiv \binom{4(a-1)}{2b+1} + 2\binom{4(a-1)}{2b-1} + \binom{4(a-1)}{2(b-1)-1} \pmod 4$$

so by the inductive hypothesis,

$$\binom{4a}{2b+1} \equiv 0 + 2 \cdot 0 + 0 \equiv 0 \pmod 4\,.$$

Next we prove (b). We have (b) is trivially true for $b = 0$ (equivalently $b = 2a$), and it is also true for $b = 1$ (equivalently for $b = 2a - 2$) as $\binom{2a}{1} = 2a$ and

$$\binom{4a}{2} = \frac{(4a)(4a-1)}{2} \equiv 2a(-1) \equiv 2a \pmod 4\,.$$

For $b = 2$ (equivalently $b = 2a - 4$) we have,

$$\binom{4a}{4} = \frac{(4a)(4a-1)(2(2a-1))(4a-3)}{4 \cdot 3 \cdot 2} \equiv (-1)(-3)(-1)a(2a-1) \equiv a(2a-1) \pmod 4$$

and
$$\binom{2a}{2} = \frac{(2a)(2a-1)}{2} = a(2a-1)\,.$$
We can prove the rest by induction on a. The base case is again easy to see. We may assume (b) is true for $a - 1$. On one hand by (2) with $c = 2b$ we have,

$$\binom{4a}{2b+1} \equiv \binom{4(a-1)}{2b} + 2\binom{4(a-1)}{2(b-1)} + \binom{4(a-1)}{2(b-2)} \pmod 4$$

and on the other hand by the recursive formula,

$$\begin{aligned}\binom{2a}{b} &= \binom{2a-1}{b} + \binom{2a-1}{b-1} \\ &= \binom{2(a-1)}{b} + 2\binom{2(a-1)}{b-1} + \binom{2(a-1)}{b-2}\end{aligned}$$

and so $\binom{4a}{2b} \equiv \binom{2a}{b} \pmod 4$ by the inductive hypothesis. □

4 Finding more kids

Let us first give some general rules. Suppose we have a protocol which works for n kids and n has a non-trivial factorization $n = ab$. Then we also have a protocol which works for a kids. We just make b copies of each of them. By symmetry, we have a protocol which works for b kids. This is called the **Factor Rule**.

Suppose we have a protocol which works for a kids and a protocol which works for b kids. We do *not* necessarily have a protocol for ab kids *unless* the probability value for both protocols are the same. Then we divide the kids into a groups of size b, use the protocol for b kids to determine which group gets the coin, and then use the protocol for a kids within the lucky group. This is called the **Product Rule**.

The restriction in **Product Rule** vanishes when $a = b$, where a common probability value is guaranteed. Thus if we have a protocol which works for a kids, then we also have a protocol for a^b kids for any b. This is called the **Power Rule**.

Let's try to apply these rules. We have already seen in section 3 that flipping the coin four times and setting $r = \sqrt{\frac{1}{2\sqrt{3}} - \frac{1}{4}}$ works for three kids.

Suppose we toss a crooked coin five times. Then

$$\begin{aligned} 1 &= \left(\frac{1}{2}-r\right)^5 + \left(\frac{1}{2}+r\right)^5 + 5\left(\left(\frac{1}{2}-r\right)^4\left(\frac{1}{2}+r\right)\right. \\ &\quad \left. + 2\left(\frac{1}{2}-r\right)^3\left(\frac{1}{2}+r\right)^2 + 2\left(\frac{1}{2}-r\right)^2\left(\frac{1}{2}+r\right)^3 + \left(\frac{1}{2}-r\right)\left(\frac{1}{2}+r\right)^4\right). \end{aligned}$$

Note that $(\frac{1}{2}-r)^5 + (\frac{1}{2}+r)^5 = \frac{1}{16} + \frac{5}{2}r^2 + 5r^4$. Setting this equal to $\frac{1}{6}$, $r = \sqrt{\frac{1}{2\sqrt{3}} - \frac{1}{4}}$ works for six kids. This is exactly the same value as the one obtained in the preceding case for three kids.

The **Product Rule** now yields a protocol for 18 kids. However, such a protocol can be derived from just the protocol for 6 kids, via a protocol for 36 kids. We apply the **Power Rule** followed by the **Factor Rule**.

Let's start by tossing the coin some number of times then look for an n that works. Let's start by tossing the crooked coins seven times. Then

$$\begin{aligned} 1 &= \left(\frac{1}{2}-r\right)^7 + \left(\frac{1}{2}-r\right)^7 \\ &\quad +7\left(\left(\frac{1}{2}-r\right)^6\left(\frac{1}{2}+r\right) + 3\left(\frac{1}{2}-r\right)^5\left(\frac{1}{2}-r\right)^2 + 5\left(\frac{1}{2}-r\right)^4\left(\frac{1}{2}-r\right)^3\right. \\ &\quad \left. +5\left(\frac{1}{2}-r\right)^3\left(\frac{1}{2}-r\right)^4 + 3\left(\frac{1}{2}-r\right)^2\left(\frac{1}{2}-r\right)^5 + \left(\frac{1}{2}-r\right)\left(\frac{1}{2}-r\right)^6\right). \end{aligned}$$

Note that $(\frac{1}{2}-r)^7 + (\frac{1}{2}+r)^7 = \frac{1}{64} + \frac{21}{16}r^2 + \frac{35}{4}r^4 + 7r^6$. Setting this equal to $\frac{1}{8}$, r is the unique positive root of $64r^6 + 80r^4 + 12r^2 - 1 = 0$. This works for eight kids.

Suppose we toss a crooked coins eight times. Then

$$\begin{aligned} 1 \;=\; & \left(\frac{1}{2}+r\right)^8 + 4\left(\frac{1}{2}+r\right)^6\left(\frac{1}{2}-r\right)^2 + 6\left(\frac{1}{2}+r\right)^4\left(\frac{1}{2}-r\right)^4 + 4\left(\frac{1}{2}+r\right)^2\left(\frac{1}{2}-r\right)^6 + \left(\frac{1}{2}-r\right)^8 \\ & + 8\left(\left(\frac{1}{2}+r\right)^7\left(\frac{1}{2}-r\right) + 3\left(\frac{1}{2}+r\right)^6\left(\frac{1}{2}-r\right)^2 + 7\left(\frac{1}{2}+r\right)^5\left(\frac{1}{2}-r\right)^3 + 8\left(\frac{1}{2}+r\right)^4\left(\frac{1}{2}-r\right)^4\right. \\ & \left. + 7\left(\frac{1}{2}+r\right)^3\left(\frac{1}{2}-r\right)^5 + 3\left(\frac{1}{2}+r\right)^2\left(\frac{1}{2}-r\right)^6 + \left(\frac{1}{2}+r\right)\left(\frac{1}{2}-r\right)^7\right). \end{aligned}$$

Note that $((\frac{1}{2}-r)^2+(\frac{1}{2}+r)^2)^4 = (\frac{1}{2}+2r^2)^4$. Setting it equal to $\frac{1}{3}$, we have $\frac{1}{2}+2r^2 = \frac{1}{\sqrt[4]{3}}$. Hence $r = \sqrt{\frac{2-\sqrt[4]{3}}{4\sqrt[4]{3}}}$ works for three kids. Setting it equal to $\frac{1}{5}$, we have $\frac{1}{2}+2r^2 = \frac{1}{\sqrt[4]{5}}$. Hence $r = \sqrt{\frac{2-\sqrt[4]{5}}{4\sqrt[4]{5}}}$ works for five kids. Setting it equal to $\frac{1}{9}$, we have $\frac{1}{2}+2r^2 = \frac{1}{\sqrt{3}}$. Hence $r = \sqrt{\frac{2\sqrt{3}-3}{12}}$ works for nine kids.

Suppose we toss a crooked coins nine times. Then

$$\begin{aligned} 1 \;=\; & \left(\frac{1}{2}-r\right)^9 + 3\left(\frac{1}{2}-r\right)^6\left(\frac{1}{2}+r\right)^3 + 3\left(\frac{1}{2}-r\right)^3\left(\frac{1}{2}+r\right)^6 + \left(\frac{1}{2}+r\right)^9 \\ & + 9\left(\left(\frac{1}{2}+r\right)^8\left(\frac{1}{2}-r\right) + 4\left(\frac{1}{2}+r\right)^7\left(\frac{1}{2}-r\right)^2 + 9\left(\frac{1}{2}+r\right)^6\left(\frac{1}{2}-r\right)^3 + 14\left(\frac{1}{2}+r\right)^5\left(\frac{1}{2}-r\right)^4\right. \\ & \left. + 14\left(\frac{1}{2}+r\right)^4\left(\frac{1}{2}-r\right)^5 + 9\left(\frac{1}{2}+r\right)^3\left(\frac{1}{2}-r\right)^6 + 4\left(\frac{1}{2}+r\right)^2\left(\frac{1}{2}-r\right)^7 + \left(\frac{1}{2}+r\right)\left(\frac{1}{2}-r\right)^8\right). \end{aligned}$$

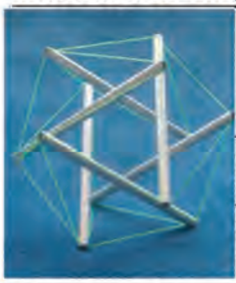
Note that $((\frac{1}{2}-r)^3+(\frac{1}{2}+r)^3)^3 = (\frac{1}{4}+3r^2)^3$. Setting this equal to $\frac{1}{4}$, $r = \sqrt{\frac{4-\sqrt[3]{4}}{12\sqrt[3]{4}}}$ works for four kids as we saw in section 2. Setting this equal to $\frac{1}{10}$, $r = \sqrt{\frac{4-\sqrt[3]{10}}{12\sqrt[3]{10}}}$ works for ten kids.

We do not have a protocol which works for seven kids. Perhaps the reader can construct one. The inspiration for this problem comes from the Hungarian Mathematical Olympiad, called the Kurschak Competition.

Chiral Icosahedral Hinge Elastegrity's Geometry of Motion

Eleftherios Pavlides, PhD AIA Roger Williams University[1], Peter Fauci, Roger Williams University
[1]epavlides@rwu.edu 401 662 7521

Introduction Hinge Elastegrity, Definitions and Transformations Presented at G4G12



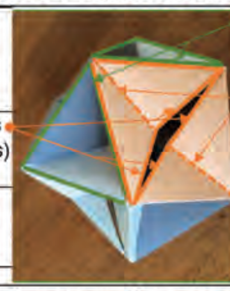
2a. 2 pairs of moving triangles frame each of 6 gates

Fig.1 (left) 6-strut 24-cable nodal tensegrity

2b. 3 orthogonal isosceles triangle faces for each of 8 irregular tetrahedra

3.a 4 free isosceles legs edges per gate (6 gates)

Fig.2 Chiral Icosahedral Hinge elastegrity rigid & moving parts

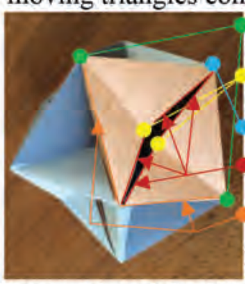


3.b 2 hypotenuse elastic hinges link each of 2 tetrahedra

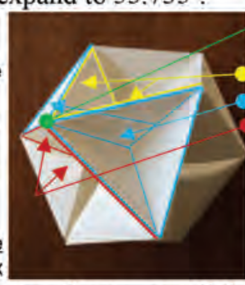
3c. isosceles leg elastic hinging pairs of right triangles forming 12 springs

Fig.3 Chiral Icosahedral Hinge elastegrity hinges & gate edges

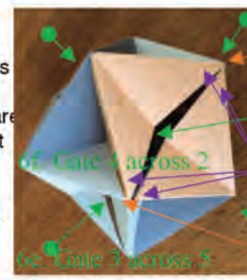
The object that gave rise to the math in this paper is "hinge elastegrities", a class of structures that originally arose from two Bauhaus exercises assigned at the Yale School of Architecture in the 1970's and investigated in a series of art projects. The key new object obtained in 1982 involved cutting slits into folded pieces of paper and weaving them into 8 irregular tetrahedra, each with 3 isosceles right-triangle faces outlining an equilateral face fig.2b The 8 tetrahedra are suspended with 12 pairs of moving isosceles-right-triangles, congruent to the tetrahedral face right triangles fig.2a giving rise to an icosahedral shape (not necessarily regular) fig.2. Each pair of right triangles is attached to each other with an elastic hinge, along one of its isosceles legs fig.2a that act as springs. The other isosceles legs are free and frame one of 6, four-sided gates, that open and close as the structure moves fig.3a. The pair of moving right triangles is also attached along their hypotenuse, with elastic hinges, to the hypotenuse of the faces of 2 tetrahedra. The 2 tetrahedra are thus linked by a hinged pair of moving triangles fig.3b. As the structure moves, the dihedral angles between the 12 pairs of moving triangles contract to 0° and expand to 180° while the 24 dihedral angles between tetrahedra and moving triangles contract to 0° and expand to 53.735°.



4a 2 top hinge pivot vertices
4b 2 bottom hinge pivot vertices
4c vertices of 45° angles between
4d free isosceles legs
4e hypotenuses share vertex hinge
fig 4 shared vertex



5a top pivot-vertex shared by 6 elements:
5b Pair of hinged triangles
5c 2 tetrahedra
5d 2 45° triangle that share vertex and frame gate at right angle to the gate framed by the pair of hinged triangles pointed with yellow arrows.



6g. Gate 5 across 3
6f. Top of hinge vertex Gate 5
6a. Gate 1 across and parallel to 6
6b. Acute (45°) angle vertices Gate 1
6c. Gate 2 across 4
6d. Top of hinge vertex Gate 3

The moving right triangles frame 3 pairs of gates, parallel to each other, that open and close in unison. Each gate has four sides, defined with four vertices: 2 bottom hinge pivot-vertices fig.4b and 2 acute isosceles angle pivot-vertices formed by the hypotenuses and the free isosceles leg of the moving triangles. fig.4e. The 12 pivot-vertices-of-acute-isosceles-angle are also the same vertices of the 12 top-of-hinge-pivot-vertices linking two adjacent moving triangles fig.4c. that frame an adjacent gate, oriented at right angle to the original gate. For example the top-of-the-hinge pivot-vertex for gate 5 fig.6g is also the pivot vertex of the acute isosceles angle for gate 1 fig.6f, which is at right angles with gate 5. The top of hinge for gate 3 fig.6d is also the pivot vertex for the acute isosceles angle of gate 1 fig.6d. Gate 5 is parallel to gate 3 and both are perpendicular to gate 1.

Each gate is framed with 2 pairs of hinged triangles fig.2a. Each tetrahedron is attached to one of the triangles of 3 hinged triangles pairs fig.7 & fig.8. The 3 elastically hinged triangle pairs, act like springs supporting and linking each tetrahedron with a pair of triangles acting as springs, to 3 tetrahedra. When the structure moves, the 3 tetrahedra rotate with opposite chirality to the original, "floating" like rigid islands in a sea of elasticity to paraphrase Julian Rimoly's of Georgia Tech, definition for tensegrities. Tensegrities are related structures made of struts, nodally connected with prestressed cables fig 1.

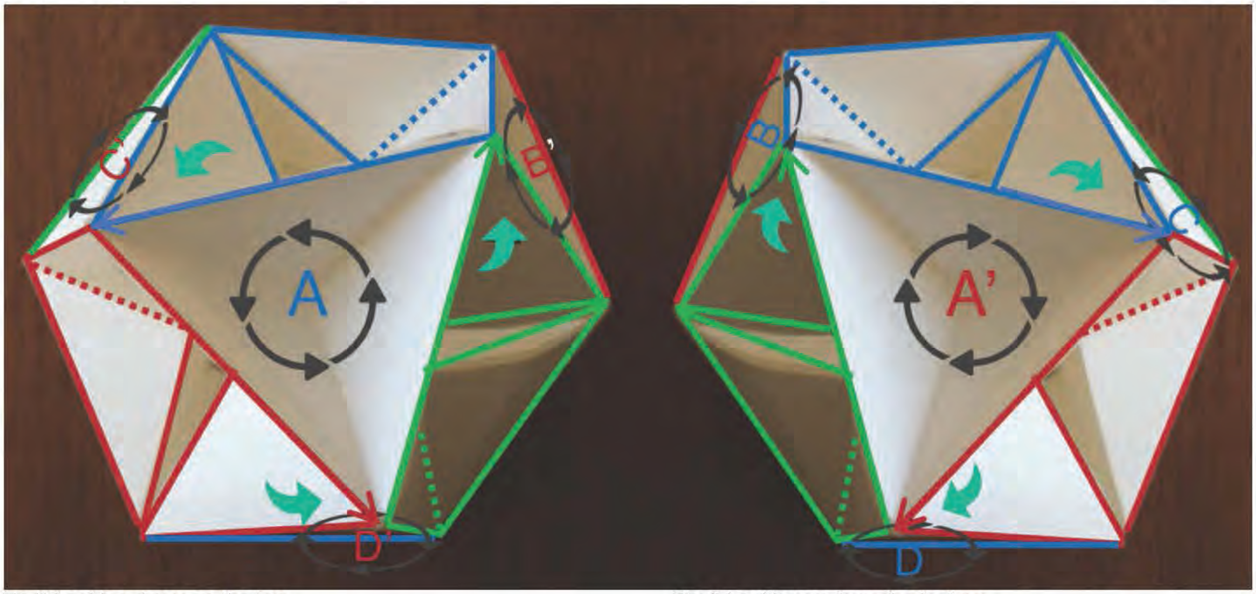


Fig.7 Looking down on the top Fig.8 Looking up from the bottom

The chiral icosahedral hinge elastegrity has noteworthy physical and geometric properties. When the 8 tetrahedra with 3 orthogonal faces, are compressed along any one of 4 axes, the hinges contract in unison, gyrating 4 tetrahedra clockwise and 4 counterclockwise, until the 8 tetrahedra rest back to back into a regular octahedron. When pulled along one of four axes, the structure extends with reverse gyration into a cuboctahedron. When external forces are removed, elastic forces in the hinges return the structure isometrically, into its original icosahedral shape (not necessarily regular). Because of similarities[i] in symmetry and elasticity of the structure with tensegrity figures Fig. 1 that maintain shape integrity by pre-stress tension alone, these new objects that maintain shape integrity through elastic hinges[ii] were named "hinge elastegrity".

The hinge elastegrity's shape-shifting through further folding was presented at G4G12 and it led to a number of familiar geometric objects, as well as some new ones fig. 9. The hinge elastegrity can flatten into a multiply covered square, morph into shapes with the vertices of each of the Platonic shapes, model the hypercube, transform into objects symmetrical to 6-strut, 12-strut, 30 strut, and 60 strut tensegrity, as well as take the shape of new figures with the vertices of congruent faces that are not regular polygonal regions. Co-presenter at G4G12, professor Thomas Banchoff, termed these figures monohedra. The object obtained through folding and trigonometry, has twelve congruent pentagons, each pentagon having one right angle. The side of the pentagon across the right angle is 0.54.. of the four equal sides. Using analytic geometry, Banchoff generalized this unique new monohedron-dodecahedron into a continuum family of monododecahedra (f=12 congruent not regular). Through folding, the smaller side can vary from expanding to be equal to the other 4 sides becoming regular dodecahedron, to decreasing to 0, the variable side becoming a point, thus the monohedron dodecahedron becoming rhombic.

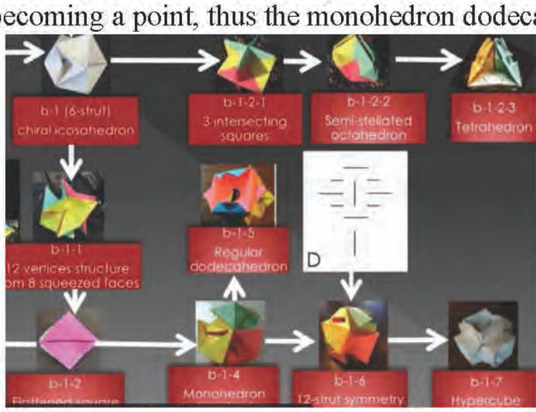


Fig. 9 Partial Elastegrity tree of forms through folding

Platonic Solids:Tetrahedron **b-1-2-3**, Cube **b-1-3**, Octahedron **b-2-2**, Regular dodecahedron **b-1-4**, Icosahedron **b-1**;

Tensegrities: Same symmetry as a 6-strut **b-1**; 12-strut **b-1-6**

Jitterbugs gyration symmetrical in 2 directions along 3 axes: Cube **b-1-3,** Flat square goes up and down, **b-1-2**;

Mono-dodecahedron continuum

Pentagonal dodecahedron 1 right angle **b-1-4**, => regular **b-1-5** => rectangle with one angle 180°

2-d Square: **b-1-2,** to **3-d** Cube **b-1-3,** to **4-d** Hybercube: **b-1-7**.

D fractal of **B** derived through 3-d folding of **b-1** into **b-1-1,** into **b-1-3** (they have the same creases but different orientation in 3-d)

Geometry of Motion of 13 axes and Movement of Vertices Outlining a Dodecahedron

The geometry of motion of the Chiral Icosahedral Hinge Elastegrity's members is presented here, in relation to 13 axes in 3 sets: a. a set of 4 tetrahedral, fig. 5, 6, 7 blue, b. a set of 3 orthogonal, fig. 5, 6, 7 yellow, and c. a set of 6 diametric axes through opposite icosahedral vertices, fig. 5, 6, 7 green.

Also we present the geometry resulting from movement of 20 vertices: 12 hinge-bottom-pivot-vertices together with 8 right-angles-vertices of the orthogonal tetrahedral faces, outline a dodecahedron. We show that a force applied on the icosahedron actuates motion of the 20 vertices replicating through movement the same geometric transformations, as those presented at G4G12 through folding: a continuum of dodecahedra shapeshifting from regular, to monohedral (congruent faces but not regular), to rhombic.

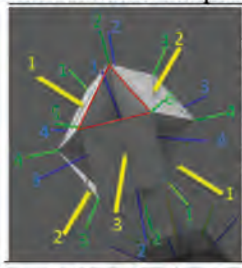


Fig.10 cuboctahedron 12 dihedral angles between moving triangles expanded to 180° forming squares; 24 dihedral angles between moving & tetrahedral triangles expanded to 54.347°

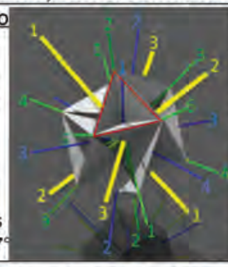


Fig.11 regular Icosahedron Dihedral angles between each of 12 pairs of moving triangles is 90°; 24 dihedral angles between tetrahedral and moving triangles is 28.72°

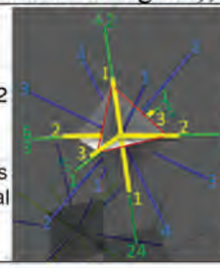
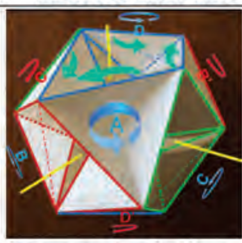


Fig.12 octahedron 12 dihedral angles between moving triangles contracted to 0°;

24 dihedral angles between moving & tetrahedral triangles contracted to 0°.

a. Tetrahedra translate and spin along 4 axes (blue)



4 tetrahedral axes through the center of opposite tetrahedra A-A', B-B', C-C', D-D'
3 orthogonal axes through the center of the gates showing,

- counterclockwise spinning of A, B, C, D
- clockwise spinning of A', B', C', D'

Tetrahedra D and B' share common pivot vertex DB' one on one side of the gate and tetrahedra A and C' share common pivot vertex AC' on the other side of the gate and rotate around their shared hinges so that common vertex C'D and AB' become congruent when the structure contracts into an octahedron.

Fig. 13 Clockwise and counterclockwise tetrahedral spinning around 4 axes as 6 gates close

When a pair of diametrically placed tetrahedra are pressed together, along any of 4 axes, defined by opposite centers of equilateral faces of the 8 tetrahedra (blue fig.5, 6, 7), the 4 axes gyrate around the structure's center fig.13. The 8 asymmetrical tetrahedra spin and slide along the gyrating axes towards the center, into an octahedron. As the structure contracts, the orthogonal faces of 2 tetrahedra sharing a pivot vertex rest back to back, squeezing between them, a folded pair of hinged moving right triangles, also hinged to the 2 tetrahedra along their hypotenuse. As the 6 gates close, the 8 right angle vertices of the orthogonal tetrahedral faces and the 12 bottom hinge pivot points, become congruent with each other and with the center of the structure.

When any two opposite tetrahedra are pulled away from each other, along any of the 4 axes, the 36 elastic hinges also actuate simultaneously movement of the entire structure. The 8 tetrahedra spin and slide away from the structure's center, in unison along the 4 axes, as the axes gyrate with reverse chirality, around the structure's center. The 12 dihedral angles of moving pairs of triangle, open to 180°, while the 24 hypotenuse dihedral angles expand to 54.375°. The 8 equilateral triangle faces of the tetrahedra rotate as they move away from the center to become the 8 equilateral faces of a cuboctahedron fig.10 red triangle.

b. Movement of the gates around 3 stable orthogonal axes (yellow)

3 orthogonal axes are defined by the centers of the 3 opposite pairs of gates. When a force actuates motion along any of the tetrahedral axes activating contraction or expansion of the 36 hinges, the 6 gates axes open and close around the 3 axes. The centers of the gates slide away in 8 directions when the structure is expanding, and towards the center when the structure is contracting.

As the dihedral angles of the moving triangles approach 0°, the angle between the free isosceles legs pivoting around the bottom-hinge-vertex also approach 0°. Simultaneously the width of the gates, the distance between the two bottom-hinge-pivot-vertices across each gate, decrease approaching zero.

As the 36 dihedral angles contract to 0° and the structure contracts into an octahedron, each of the 6 sets of 4 rotated isosceles legs, edging the gates, become congruent with the axes of the octahedron. The orthogonal axes of the structure become congruent with the orthogonal axes of opposite vertices of the octahedron. In addition a) the 2 hinges linking the 4 triangles surrounding each gate, b) the 4 edges of the 2 pairs of tetrahedra that each shares the top of the 2 hinge vertices of a gate on either side fig.4a. as well as c) the 4 gate-edges contracted to 0°, all 10 (2 hinges+4 tetrahedral edges+4gate-edges) become congruent.

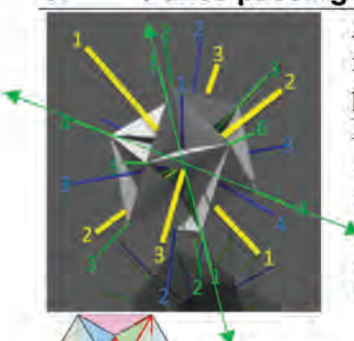
As tension is applied on any of the 4 tetrahedral axes, the gates open pivoting with opposite chirality around the 3 axes. As the dihedral angles of the moving triangles open towards 180°, the gates' width

decreases again. When the dihedral angle becomes 180°, the two moving right triangles on either side of the hinges, align into 2 bigger right triangle, on either side of the gate, flattening into a square Fig.14. The free isosceles legs that are the gate edges, rotate 180°, closing the gates, and forming the diagonal of the square faces of the cuboctahedron.

Fig.14 Diagonals of cuboctahedron square faces show closed gates when structure is expanded.

The 3 orthogonal axes pass through the centers of the 6 diagonals of the cuboctahedron squares that are the closed gates of the structure. Somewhere between the gate-edges closing by rotating the angle between them to 0°, when they align becoming congruent with the orthogonal axes and each other, and closing by rotating to 180°, expanding into the diagonals cuboctahedron square faces, the opening of the gates' becomes maximum. As we will discuss with more detail below the maximum gate width is achieved when the dihedral angles between moving triangles is 90°.

c. 6 axes passing through diametrically opposite icosahedral vertices



A set of 6 axes (green) is defined by pairs of diametrically opposite vertices of the icosahedron, forming 3 pairs of X. Each X passes through the 4 acute-angle-pivot-vertices of two parallel gates on either sides of the structure's center Fig.4b. In the illustration Fig.15 the 3 pairs of X are formed by axes 4 & 2, 3 & 6, 5 & 1. As the structure contracts the 3 pairs of X pivot towards each other, as they close around the 3 orthogonal axes, until each pair of axes becomes congruent with the orthogonal axes that pass through the gate centers 4 & 2 with axis 1, 3 & 6 with axis 2, and 5 & 1 with axis 3.

Fig 15 Diametric axes 4 & 2, 1 & 3, 5 & 1 form X, congruent with 3 orthogonal axes when closed.

When the structure expands into a cuboctahedron, each pair of axes passing through diametrically opposite vertices forming an X open to 60°.

Fig 16 When structure expands into a cuboctahedron 3 X each open to 60°

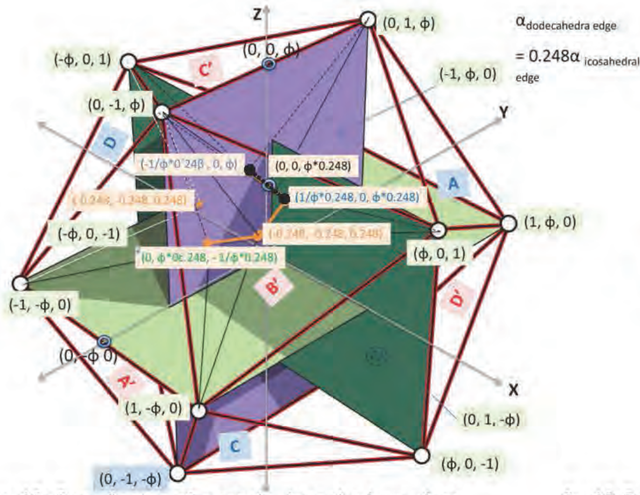
d. Regular dodecahedron, to monohedron, to rhombic dodecahedron through movement

Examining the movement of the 13 axes, and in particular pondering the closing, opening, and closing again of the gates, as the structure contacts into an octahedron and expands into a cuboctahedron, raised the question, at what point is the gate width maximal?

Fig.17 dihedral angle = 90°

Attempting to answer the question of when do the gates achieve maximal width led to the realization that when the distance between 2 vertices across a gate is equal (fig.17 red line) then the implied 12 tetrahedra, created by drawing the 6 red lines, are congruent with the 8 tetrahedra suspended by the moving triangles (fig.17 green lines) and therefore the dihedral angle between each pair of moving triangles is 90°. The distance from each of the 20 right angle vertices of the orthogonal faces of the 20 congruent tetrahedra to the 3 adjacent right angle vertices of the 20 congruent tetrahedra is equal and the 20 vertices outline a regular dodecahedron. 2 of the vertices shown with black dots fig.18 show the width of a gate, which is equal to the 30 edges the regular dodecahedron shown on fig.19. In fig.18 the 8 "floating" tetrahedra are A, B, C, D and diametrically opposite with reverse chirality A', B', C', D'. In fig.19 the bottom of hinge pivot vertices are indicated with the letter of the two tetrahedra sharing the hinge: AB', AC', AD', BC', BD', CD', A'B, A'C, A'D, B'C, B'D, C'D, the right angle vertices of the orthogonal faces of the 8 floating tetrahedra are indicated with A, B, C, D, A', B', C', D'

$a_{dodecahedron} = r_{u\ dodecadron}$ (radius of a circumscribed sphere in a dodecahedron) $/\ 1.401258538.. \Rightarrow a_{dod} = 0.248 a_{ico}$ (derivation below)



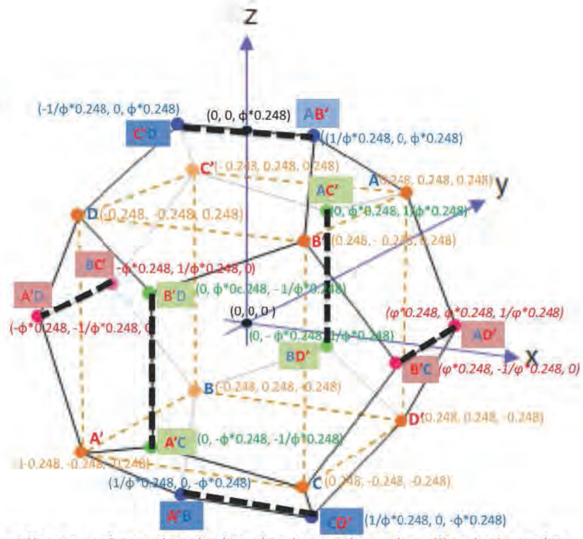


fig.18 Coordinates of a regular icosahedron when The dihedral angles are 90°

fig.19 Coordinates of regular dodecahedron when the dihedral angles are 90°, pentagonal sides are equal, doted lines indicted gates.

$r_{u\ dod} = r_{i\ ico(\text{radius of inscribed icosahedron}\ \iota)} - h_{\text{height of tetrahedron}} \Rightarrow 0.7558a_{icosa} - 0.408a_{icosa} = 0.348a_{icosa}$

- because $r_{iicosahedron(\text{radius of inscribed icosahedron}\ \iota)} = 0.7557623142a_{icosa}$
- and $h_{tetrahedron} = r_{iocta\ (\text{radius of inscribed octahedron})} = r_{ioctacedron} = 0.408a_{iocta} = 0.408a_{iico}$ ($a_{octa}=a_{icosa}$)

so $r_{u\ dod}$ / 1.401258538 = $\boxed{0.248a_{icosahedron}}$ is the width of the gate when the icosahedron is regular.

1. When a force contracts the icosahedron, so that the moving triangle hinge dihedral angles is less than 90°, then the 20-vertice-dodecahedron decreases isometrically and proportionally, and remaining regular, until the 20 vertices become congruent with the center of the structure, and shrink to a point when the structure contracts into an octahedron.
2. When a force expands the icosahedron so that the moving triangle dihedral angles are greater than 90°, then the width of the 6 gates decrease again while the other 4 pentagonal sides increase.

Therefore 6 gate Max width = $0.248a_{ico}$ when the dihedral angles between moving triangles = 90°

- A polyhedron with congruent non-regular faces, (4 equal pentagonal sides but one side smaller) has been termed a monododecahedron, in G4G12. With professor Banchoff we presented a similar transformation of the chiral icosahedral hinge elastegrity through folding giving rise to a continuum of pentagonal monododecahedra, until the decreasing side = 0, when the dodecahedron is transformed to rhombic.
- When the dihedral angle expands to 180 ° the gate width is 0, the 2 bottom of the hinge vertices on either side of each gate become congruent, reducing the number of vertices from 12 to 6. Together with the 8 right angle vertices of the orthogonal faces of the tetrahedra, the dodecahedron is a 14 vertex rhombic.

From a physics point of view and applications (experimentally derived)

- When the stiffness of a shape-memory membrane creates an elastegrity with the dihedral angles sagging below 90° (or when the structure starts contracted with 0° dihedral angles), a matrix of chiral icosahedral hinge elastegrities behaves as a tensile spring.
- When the a shape-memory membrane is stiff and the 12 dihedral angles are greater than 90° (or start expanded to 180°), a matrix of chiral icosahedral hinge elastegrities acts as a compression spring.

There are several possible applications for the Chiral Icosahedral Hinge Elastegrity cited in a footnote in the paper for G4G12, including a number of existing tensegrity applications that may be improved with the additional properties of hinge elastegrities and some novel applications specific to the additional unique properties of hinge elastegrities.

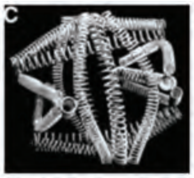


Fig. 20 Ingber Model

One of the applications proposed for tensegrities, is Donald Ingber's, conjecture that all biological structure is hierarchically ordered tensegrities. Ingber is a founding director of the Wyss Institute for biological engineering and has been publishing on this topic for over 35 years. Ingber proposed a model of tensegrities where compression coil springs take the place of struts and tensile coil springs take the place of cables. This model can be also seen in his most recent publication *Multi-scale modeling reveals use of hierarchical tensegrity principles at the molecular, multi-molecular, and cellular levels* C. Reilly, D. Ingber, Extreme Mechanics Letters 20, (2018) 21. In this most recent article Ingber proposes a force and energy distribution

Fig.21 a. & b. dihedral angle>90° b. dihedral angle<90° c. NPR (Negative Poisson's Ration gets narrower as it gets shorter)

argument. Hinge elastegrities address several questions that tensegrities leave unanswered including accounting for the fact that numerous independent biological papers expressed astonishment in measuring experimentally NPR in different parts of the anatomy of various species, suggesting that Negative Poisson's Ratio is ubiquitous in all of the architecture of life. For links to these articles please email to request. Additional elastegrities suggest the design of pumps for non-Newtonian fluids, that are prevalent in biological structure, and self-assemble into a structures at hierarchically different scales with smaller scale elastegrity components through folding a shape memory membrane.

If you are interested to collaborate in investigating the force distribution and energy transmission of a matrix of Chiral Icosahedral Hinge Elastegrities fig.21 please conduct me at epavlides@RWU.edu. It may open the gate to numerous applications that you may be interested to develop collaboratively.

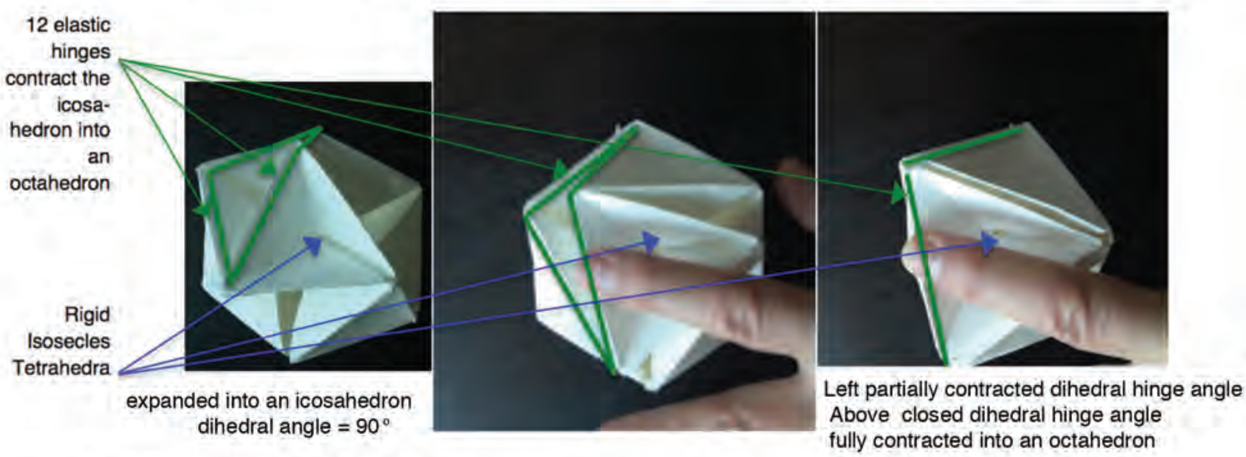


Fig.22 Chiral Icosahedral Hinge Elastegrity contracting into an octahedron

---

[i] Structurally both tensegrities and elastegrities are networks of rigid (linear struts for tensegrities, irregular tetrahedra for the icosahedral hinge elastegrity) and elastic (pre-stressed cables acting in pure tension for tensegrity and elastic hinges for elastegrities)

[ii] The term "hinge" differentiates elastegrities from those that may be termed "nodal". Nodal elastegrities can be created by replacing tensegrity's pre-stressed cables, holding together struts with springs. Additional hinge elastegrities have been created from further folding, weaving, and inverting the original icosahedron. One may consider elastegrities both "nodal" and "hinge" as the general family of structures that tensegrities are a special subcase

Life Algorithms

Tomas Rokicki

June 28, 2018

Abstract

We describe some ideas on how to implement Conway's Game of Life.

1 Introduction

Conway's Game of Life serves as the "Hello World" of recreational computing. Many of us who cut our teeth on minicomputers or microcomputers will remember writing our own implementations, and then struggling to make them faster [5]. In this paper we briefly describe some of the more interesting ways to implement Life, without taking the fun out of getting all the details of implementation exactly correct.

We will focus on calculating subsequent generations from current generations, ignoring the impact on other operations such as loading and saving patterns, setting and clearing cells, calculating the current population of the universe, and displaying the pattern. We will also ignore the impact of the borders (if any) of the universe. In a real Life program, all of these are important aspects that should be considered before adopting a too-complex algorithm.

All the code for this paper is available at https://github.com/rokicki/lifealg along with more information and performance comparisons. We include the algorithms in the Golly program as well as high-performance algorithms created by Adam Goucher as part of his lifelib project.

2 Elementary Approaches: Arrays

2.1 Basic Algorithm

The most fundamental implementation of the Game of Life uses two two-dimensional arrays representing the current universe and the universe in the next generation. Computing the next generation array from the current generation array involves iterating over all the cell positions, calculating the number of live neighbors to that cell from the current generation array, and then determining if the next generation cell is alive or dead based on this neighbor count and the life status of the current generation cell. After this calculation, we either swap the current and next generation arrays, or copy the next generation array into the current generation array.

```
void nextgen(unsigned char u0[][W], unsigned char u1[][W]) {
   for (int i=1; i+1<H; i++)
      for (int j=1; j+1<W; j++) {
         int n = u0[i-1][j-1] + u0[i-1][j] + u0[i-1][j+1] + u0[i][j-1] + u0[i][j+1] +
                 u0[i+1][j-1] + u0[i+1][j] + u0[i+1][j+1] ;
         u1[i][j] = (n == 3 || (n == 2 && u0[i][j])) ;
      }
}
```

In some sense, this is the best we can do; for a completely random pattern that fills our array, we can only make constant-time improvements over this basic algorithm when calculating a single next-generation step. But these constant-time improvements can be significant. Furthermore, we can exploit

the fact that subsequent generations are not random, but show certain behaviors; for instance, when starting from a totally random pattern, the average population density decreases quickly, and regions of the space achieve some degree of stability over limited time intervals.

Also, once we have played with many random universes, our attention may shift to specific constructions that exhibit specific behavior, such as puffer trains or patterns that allow you to play Tetris. These patterns are frequently sparse and exhibit a high degree of regularity that can be exploited for speed.

2.2 Simple Improvements

The previously given simple algorithm runs at about 270 million cell generations per second on a laptop, so it is easily fast enough to animate a large 2,000 by 2,000 universe at more than 60 generations per second. If all you want is a pretty display, there's likely to be no reason to do anything more complicated. But let's assume you are interested in much larger patterns, or in running them far more quickly.

Some of the changes we will make will have much larger impacts than we might expect. For instance, replacing the line

```
u1[i][j] = (n == 3 || (n == 2 && u0[i][j])) ;
```

with a precalculated table given the value for the next generation and neighbor count, as in

```
u1[i][j] = nextval[u0[i][j]][n]
```

can nearly double the speed to 470M cell generations per second, even though the total number of instructions executed per generation in the two versions is almost identical. Modern CPUs predict conditional expressions in order to evaluate multiple instructions per cycle, and they pay a significant performance penalty when such branches are not easily predictable. Replacing conditionals with lookups into small tables and straight line code can help the processor run faster.

2.3 Integration

Instead of checking all eight neighbors for each cell, we can instead accumulate cell accounts for each row and then across rows, with code like the following:

```
void nextgen(unsigned char u0[][W], unsigned char u1[][W+1]) {
   for (int i=0; i<=H; i++)
      u1[i][0] = 0 ;
   for (int j=0; j<=W; j++)
      u1[0][j] = 0 ;
   for (int i=0; i<H; i++)
      for (int j=0; j<W; j++)
         u1[i+1][j+1] = u0[i][j] + u1[i][j+1] + u1[i+1][j] - u1[i][j] ;
   for (int i=1; i+1<H; i++)
      for (int j=1; j+1<W; j++) {
         unsigned char n = u1[i+2][j+2] - u1[i+2][j-1] - u1[i-1][j+2] + u1[i-1][j-1] ;
         u0[i][j] = (n == 3 || (n == 4 && u0[i][j]))
      }
}
```

This algorithm uses `u1` to integrate the original universe in two dimensions (using an unsigned data type to avoid undefined behavior), and then calculates the new result back into the `u0` array. This is just about as fast as the original algorithm for the normal Conway game of life, but it is significantly faster for a variant called "Larger than Life", where we use a larger neighborhood. With this algorithm we can compute "Larger than Life" for a large neighborhood (such as 201x201) about as fast as we can calculate the normal Game of Life.

2.4 Single Instruction Multiple Data

In the simple algorithm presented, we use two arrays, each with a full byte (eight bits) per cell; this wastes memory, since each cell has only two potential values. But it also wastes computation. Our neighbor count is a full integer value, with 32 bits, but the largest value it will ever need to store is 8, so can get away with four bits. Modern CPUs operate on 64-bit wide words (or even wider; 128-bit and 256-bit one-cycle operations are now common, and 512-bit one-cycle operations are starting to appear). We can use this to speed up our calculations by using a wider type for our arrays and packing more than one cell state in this wider type.

Let us say we use 64-bit unsigned words, and we allocate four bits for each cell; this lets us put 16 cell states in a single 64-bit word. We will assign 16 horizontally-adjacent cells to each 64-bit word. We are now wasting 3 bits for each cell—but we do this so we can accumulate neighbor counts for all 16 cells in one sequence of instructions by using carryless addition.

Consider adding the decimal values 314 and 271; as long as the sum of digits in each position is never greater than 9, we will never generate a carry. For hexadecimal, we can parallelize small additions up to a maximum value of 15 as long as we ensure no small addition ever exceeds 15. Since our maximum neighbor count is only 8, this is guaranteed.

Once we have the 16 neighbor sums packed into a single 64-bit word, we need to calculate the 16 resulting cells. Using a lookup table to do this would require 16 separate shifts, masks, lookups, and adds, but some clever bit manipulation and masks can do the work for us. In addition, it is easier to calculate the sum of all nine cells and then modify our "stay alive" counts slightly. Getting the shifts and masks exactly correct is a bit tedious but not extremely so; the resulting code looks like this:

```
void nextgen(unsigned long long u0[][W], unsigned long long u1[][W]) {
   for (int i=1; i+1<N; i++)
      for (int j=0; j<W; j++) {
         unsigned long long pw = u0[i-1][j] ;
         unsigned long long cw = u0[i][j] ;
         unsigned long long nw = u0[i+1][j] ;
         unsigned long long n = (pw << 4) + pw + (pw >> 4) + (cw << 4) +
            (cw >> 4) + (nw << 4) + nw + (nw >> 4) ;
         if (j > 0)
            n += (u0[i-1][j-1] + u0[i][j-1] + u0[i+1][j-1]) >> 60 ;
         if (j+1 < W)
            n += (u0[i-1][j+1] + u0[i][j+1] + u0[i+1][j+1]) << 60 ;
         unsigned long long ng = n | cw ;
         u1[i][j] = ng & (ng >> 1) & (~((ng >> 2) | (ng >> 3)))
                                              & 0x1111111111111111LL ;
      }
}
```

This code runs at about 6.1 billion cell generations per second, about thirteen times faster than the previous code. An excellent book on such bit tricks is [7].

But that is not nearly as fast as we can make it.

2.5 Bitwise Parallel Addition

In order to do carryless addition, we packed 16 cell values into a 64-bit integer, because we needed four bits in the neighbor sum to represent values up to 9. Instead of storing those four bits in a single register, we can store the bits in separate registers. We can use a single register to store the low-order bit for 64 values, a second register to store the next higher-order bit for 64, values, and so on. This requiring four registers to represent 64 4-bit neighbor counts. With this representation, we can store 64 Life cells in a single 64-bit integer.

Internally the CPU performs arithmetic addition using logical operations. To calculate the low-order bit of the sum of two bits (known as a half adder), we calculate the exclusive-or of the two bits;

to calculate the high-order bit, we calculate the 'and' function. To do this for three bits, we cascade two half-adders and add an additional 'or' function to merge together the carries from both sums. Processors contain bitwise logic operations that allow us to perform these sorts of calculations on the entire width of a register in a single instruction. So to calculate the low and high order bits of the bitwise sums of each bit in a 64-bit register, we can use the following code:

```
void halfadder(unsigned long long a, unsigned long long b,
               unsigned long long &s0, unsigned long long &s1) {
   s0 = a ^ b ;
   s1 = a & b ;
}
void fulladder(unsigned long long a, unsigned long long b, unsigned long long c,
               unsigned long long &s0, unsigned long long &s1) {
   unsigned long long t0, t1, t2 ;
   halfadder(a, b, t0, t1) ;
   halfadder(t0, c, s0, t2) ;
   s1 = t1 | t2 ;
}
```

To sum more than three bits, we cascade full adders as needed.

There are a few small optimizations we can perform. We don't need the highest-order bit, since total cell counts of 8 and 9 have the same effect as total cell counts of 0 and 1. Overall, using 64-bit registers and standard C++ code, we can achieve 13.6 billion cell generations per second using this approach.

In some environments, it may be useful to work on full bitplanes instead of registers. On modern CPUs this won't give maximum performance because it requires too much memory bandwidth. But on the HP48 calculator, for instance, where bitwise logical operations on bitplanes are supported, the following code executes Life quickly even in interpreted RPL:

```
GEN1 << {#0 #1} DUP2 SH OVER LX ROT REVLIST
  SWAP OVER SH 5 ROLLD 4 ROLLD SH 4 PICK LX
  ROT 3 PICK + NEG + 4 ROLLD NEG + LX + NEG >>
SH << DUP2 OVER DUP ROT {#FFFh #FFFh} SUB
  LX 3 DUPN 7 ROLLD GXOR 5 ROLLD GXOR + >>
LX << {#0 #0} SWAP GXOR >>
```

2.6 Minimum Logical Operation Count

There have been some published papers that incorporate parallel bit-wise addition for calculating Life, often focusing on performance. The ones I have seen use suboptimal sets of logical operations. To provide a basis for comparison, we will imagine that we have an 8x8 section of the universe in row-major order in a single 64-bit word, and we wish to compute the inner 6x6 section in as few CPU instructions as possible. The principles we use apply to many conventional universe layouts but focusing on a square in a single word and ignoring the edges allows us to simplify our presentation without losing generality.

Just cascading half-adders to generate the low-order three bits of the sum of neighbors would look something like this:

```
lifeword gen4(lifeword a) {
   lifeword a0, a1, b0, b1, c0, c1, d1, d2, e1 ;
   add3(a>>9, a>>8, a>>7, a0, a1) ;
   add3(a<<9, a<<8, a<<7, b0, b1) ;
   add3(a<<1, a>>1, a0, c0, c1) ;
   add3(a1, b1, c1, d1, d2) ;
   e1 = c0 & b0 ;
   c0 ^= b0 ;
```

```
   d2 ^= e1 & d1 ;
   d1 ^= e1 ;
   return d1 & (~d2) & (c0 | a) ;
}
```

This takes 29 Boolean operations and 8 shifts. But since the operations are bitwise, the first two full adds are computing the same result, just shifted. We can do a horizontal sum of three bits and then use that result three times with the following code (which sums all nine cells, not just the eight neighbors). When applied to dense bitmaps, this is equivalent to retaining the horizontal sum for each row across the three rows that require it.

```
lifeword gen4(lifeword a) {
   lifeword a0, a1, b0, b1, c1, c2 ;
   add3(a>>1, a, a<<1, a0, a1) ;
   add3(a0>>8, a0, a0<<8, b0, b1) ;
   add3(a1>>8, a1, a1<<8, c1, c2) ;
   c2 ^= b1 & c1 ;
   b1 ^= c1 ;
   return (b0 ^ c2) & (b1 ^ c2) & (a | b0) ;
}
```

This takes only 23 Boolean operations and six shifts. But we can improve this further. The first full-add is composed of two half-adds, and instead of throwing away the sum of the first two elements, we can preserve those and use this to compute a proper sum of neighbors rather than a sum of all elements. Further, rather than compute three binary bits of the sum, we can use a simpler expression that determines if precisely one of four bits is set. The final code looks like this (after expanding all full adds):

```
lifeword gen3(lifeword a) {
   lifeword aw = a << 1, ae = a >> 1,
       s0 = aw ^ ae, s1 = aw & ae,
       hs0 = s0 ^ a,
       hs1 = (s0 & a) | s1,
       hs0w8 = hs0 >> 8, hs0e8 = hs0 << 8,
       hs1w8 = hs1 >> 8, hs1e8 = hs1 << 8,
       ts0 = hs0w8 ^ hs0e8,
       ts1 = (hs0w8 & hs0e8) | (ts0 & s0);
   return (hs1w8 ^ hs1e8 ^ ts1 ^ s1) &
          ((hs1w8 | hs1e8) ^ (ts1 | s1)) & ((ts0 ^ s0) | a);
}
```

This code requires only 19 Boolean operations and 6 shifts. I have not proved this to be optimal but I know of no better solution.

2.7 SSE and AVX2

Processor designers have been incorporating a simple vector processor in the form of wider registers and explicit SIMD instructions for many years. Using these facilities allows us to extend our techniques above from the default register width of 64 bits to 128 bits (using SSE), 256 bits (using AVX2), and even 512 bits (using AVX-512); in each case we potentially gain another doubling of performance. For SSE on my laptop, I was able to achieve 23.9 billion cell generations per second, nearly 100 times faster than our original simple algorithm. While in some cases carefully written C++ code can be automatically translated into these vector operations by the advanced compilers of today, typically to get maximum performance the programmer must resort to the very careful use of low-level "intrinsic" functions. The principles are the same as described previously, but instead of using basic C++ operations on 64-bit words, special intrinsic functions are called on special SSE datatypes. Describing how to write this code is beyond the scope of this paper, but simple examples are given in the code repository referenced above.

2.8 Multithreading

More performance is still available to us through the use of multithreading. Modern CPUs have multiple cores, with two and four being common on laptops, four and eight common on desktops, and many more frequently available on even small server computers. It is simple to assign each core a region of the life universe to work on. On my two-core laptop using four threads allowed me to nearly double the performance to 45.7 billion cell generations per second. On a 4-core desktop CPU, performance is 130 billion cell generations per second; on an 8-core server CPU, performance is 230 billion cell generations per second.

At this calculation speed we are nearing the memory bandwidth of modern CPUs, and additional tricks to exploit caching and memory locality may need to be used. For instance, instead of advancing the whole field one generation, we can advance smaller subregions that fit in the cache multiple generations at once. As AVX instructions get wider, this may be of increasing importance.

2.9 Graphics Processing Units

The performance of current CPUs is dwarfed by the performance of modern graphics processing units (GPUs), which frequently contain thousands of smaller, slower processing units that all work simultaneously. On these devices, the bit tricks that worked so well on CPUs become even more effective, but transferring the field data to and from the different cores of the GPU can be a challenge. In addition, exploiting GPUs can be extremely tedious and challenging, requiring knowledge of APIs such as CUDA or OpenGL as well as architectural knowledge of the specific graphics card to be used. Nonetheless, performance gains of between one and two orders of magnitude can be obtained using common gaming cards [1].

3 Work Smarter, Not Harder

So far we have focused on speeding up the low-level calculation of the next generation, without any consideration for exploiting empty space or regularity in the pattern and its evolution. While these low-level tricks are useful, most performance improvements in simulating Life on interesting patterns are obtained by using smarter algorithms rather than making the bit banging faster.

Our algorithms so far have been for small universes limited by the size of a 2D array we can fit in memory. Our algorithms from here on out will support unbounded universes, limited only by the data type used for coordinate arithmetic. These algorithms are all designed around a data type or types used as a container of a different data type used as a node. For containers, we may use lists, hash tables, or trees; the nodes can be individual life cells, or they can be smaller subfields within which algorithms from the previous section are used.

3.1 Associative Containers

One of the simplest implementations of Life is to just use a container of the coordinates of live cells; we can use a dictionary in Python, or a hash in Lisp, but we will describe using a set in C++. The code is just:

```
void nextgen(const univ &src, univ &dst) {
   map<pair<int, int>, int> ncnt ;
   for (auto i=src.begin(); i!=src.end(); i++) {
      ncnt[make_pair(i->first-1, i->second-1)]++ ;
      ncnt[make_pair(i->first-1, i->second)]++ ;
      ncnt[make_pair(i->first-1, i->second+1)]++ ;
      ncnt[make_pair(i->first, i->second-1)]++ ;
      ncnt[make_pair(i->first, i->second+1)]++ ;
      ncnt[make_pair(i->first+1, i->second-1)]++ ;
      ncnt[make_pair(i->first+1, i->second)]++ ;
      ncnt[make_pair(i->first+1, i->second+1)]++ ;
```

```
   }
   dst.clear() ;
   for (auto i=ncnt.begin(); i!=ncnt.end(); i++)
      if (i->second==3 || (i->second == 2 && src.find(i->first) != src.end()))
         dst.insert(i->first) ;
}
```

In this code we use a `map` to calculate a neighbor count for each cell, and then use that to calculate the new generation counts. This code only calculates at about 500,000 cell generations per second, but—unlike all previous code—it executes in time roughly linear in the number of live cells, rather than the total size of the universe. (There's a log factor due to the use of `map` and `set` that I am ignoring.) In addition, there are no bounds on the size of the universe except for the maximum coordinate that fits in an int. Despite all the advanced bit-level machinery we put in place previously to push the cell generations per second as high as possible, for many uses, the simple algorithm just given is more useful. Realistic interesting patterns tend to be sparse and non-rectangular; doing fast bit computations on, or even just examining, empty space is extremely wasteful. For example, using our fastest AVX2 algorithm on a 4K x 4K universe on the `lidka` Methuselah takes 0.223 seconds to get to generation 512; using the sample algorithm above, with all the map overhead, accomplishes this in 0.0014 seconds.

Actual performance of these algorithms is harder to pin down than for the brute force algorithms given previously because they are highly pattern-dependent. Some optimizations work well for some patterns but fail for others.

3.2 Fat Nodes

The algorithm just given was slow in terms of cell generations per second, but still performs well for many uses because it focuses only on the actual live cells of the universe. Many interesting patterns do not fit perfectly into a nice bounding rectangle or are extremely sparse. For instance, the recently found Gemini oblique spaceship fits in a 217,807 by 4,220,191 bounding box but has a population of only 846,278; fewer than one cell in a million is alive. Even at 230 billion cell generations per second using eight cores in a server with the required 230GB of memory with a flat bitplane representation, it would take at least 4 seconds to calculate each generation; the simple algorithm shown above on a single core easily achieves nine generations a second using only 60MB of memory.

In practice, we want to strike a balance between time spent managing the shape of the active region, managing knowledge of what is changing and what is stable, and computing the low-level next generations for some cells. The easiest way we do this is to use fatter nodes—nodes that contain more than a single cell of the universe. We store these nodes in some sort of container structure, and calculate the next generation by scanning this container, finding the relevant fat nodes and their neighboring nodes, and computing the next generation of each node. If the result has any cells set, we create a new node in the next generation containing those cells.

Within a fat node, to compute the next generation quickly, we can use all the fancy bit manipulation we described in the previous section of the paper. This can give us a speedup of more than 100. We will want to carefully set the size of the fat nodes, trading off efficiency within a fat node against the wasted computation on areas within the fat nodes that do not contain any active cells.

In some cases it may be simpler to have the fat nodes overlap, so some cells may be represented by bits in more than one fat node. This eliminates the necessity to consider the neighbors when calculating the next generation, but requires us to update more than one fat node with the results.

3.3 Shifting Coordinate Systems

A complication for fat nodes is that they have nine neighbors; that is a lot of neighbors to manage. A common trick that is used to improve performance is to use a shifting coordinate system. For instance, if I'm keeping two generations, I may have a node maintain the rectangle $(0,0)-(15,15)$ for the even generation but $(1,1)-(16,16)$ for the odd generation. If I do this then I only need to examine three neighbors for each calculation; I can calculate the state of $(1,1)-(16,16)$ from the prior states of

$(0,0)-(15,15)$ from the current cell, $(16,0)-(17,15)$ from the node to the right, $(0,16)-(15,17)$ from the node down, and $(16,17)-(16,17)$ from the node down and to the right. This approach can simplify calculation code significantly.

In some cases it may be simpler to shift the coordinate system shift continuously, so a cell keeps information about a moving rectangle of space over time; for instance, a given cell might track the region $(g,g)-(g+15,g+15)$ in generation g.

Another way to reduce the count of neighbors to examine is to arrange the fat nodes in a brick-laying pattern rather than a simple lattice; this reduces the neighbor count from eight other fat nodes to only six.

3.4 List Containers

The simplest container to consider is a simple list (possibly implemented with an array). Each list element could contain the two coordinate values of the fat node and the cells of the fat node itself. Other state information can be added if appropriate. If the list is organized in raster order, it can be scanned by multiple simultaneous pointers that track corresponding neighbor values in the rows above and below, eliminating the need for a hash table or other mechanism to look up the relevant fat nodes.

In our shootout below we compare list containers containing single cells, 8x8 fat nodes, and 16x16 fat nodes.

3.5 Tree Containers

Trees work well as containers because there's no need to store coordinate, neighbor, or parent information; all of that can be passed as parameters down the call stack. Furthermore, we can maintain a hierarchy of population and status information. With this organization we don't even need to examine large regions of stable nodes to see if we can skip them; we can skip an entire region of any size once we reach the appropriate tree level.

By passing neighbor information down the call stack, we do not need an associative container to randomly locate neighbors; we are always directly passing the appropriate neighbors from the appropriate tree calls. Alternatively, we do not need to maintain neighbor pointers.

Finally, trees can be very efficient in memory consumption because only the leaves have actual data, and we can size the leaves so the bulk of memory is consumed by the leaves, while still supporting appropriate status information higher up in the tree.

In our shootout we find that simple list containers tend to outperform tree containers, but if we add support for skipping stable regions, tree containers often regain the lead.

3.6 Stable Regions

In the game of life, large random regions eventually settle down, usually to a constellation of stable lifeforms and lifeforms of period 2, with the occasional larger-period oscillator and some gliders escaping the region. By storing two consecutive generations in each fat node, along with a flag indicating whether that region is stable (i.e., either not changing, or only changing with a period of 2), we can quickly skip over sections of space that are not changing.

Some care must be given to neighboring regions. If a neighboring region is changing, we may need to recompute an otherwise stable region. We can optimize this by keeping flags on stability for the border of the node as well as the node as a whole. Only if the adjacent border is changing do we need to recompute a neighboring node.

When using list containers, skipping large stable or period-2 regions can become a bottleneck. Using a tree container instead can resolve this. Another alternative is to separate active and idle fat nodes into their own lists.

3.7 Hashlife

A simple breeder defeats all the algorithms we've described so far. For this pattern, space increases at $O(n^2)$ and computation time increases at $O(n^3)$, so computing the pattern at generation $2n$ takes eight

times as much time as computing the pattern at generation n. The majority of the pattern becomes waves of gliders, which our algorithms so far do not optimize.

William Gosper introduced the amazing hashlife algorithm in 1983 [3], though it was not widely used until Golly was released in 2005. Hashlife represents the universe using a quadtree representation where identical subtrees are coalesced or canonicalized; this provides dramatic compression of space for many patterns. The algorithm caches the result of computing Life for any particular node in the quadtree to reuse this computation across generations. For more information see [6].

The real magic of hashlife, however, is that it does not just calculate the next generation for each tree node; instead, it jumps further ahead in the future than that. For a quadtree node at level n, representing a 2^n x 2^n region of space, we calculate the result at 2^{n-2} generations forward. This simplifies our recursive calls and allows hashlife to compute patterns many billions of generations in the future.

For most Life patterns that people run, hashlife is astonishingly faster than anything we have presented so far. It is not uncommon for a pattern to run somewhat normally for a second or two but suddenly explode in speed, generating trillions of generations in seconds, and often unexpectedly. But for patterns that stay essentially random for long periods of time, or have significant amounts of entropy, hashlife can be much slower, although it is highly unusual for hashlife to be significantly slower than any non-hashlife algorithm over several minutes of runtime.

Hashlife is a fascinating algorithm. During its execution the only thing it is doing is hashtable lookups and hashtable insertions (and the occasional garbage collection); almost no time is spent doing a primitive cell computation (at least not in the standard game of life). The performance of hashlife is entirely dependent on the efficiency of the hashtable implementation used.

Hashlife is reputed to take a tremendous amount of memory, and this can certainly happen. But a "tremendous" amount of memory is a changing quantity. That 32MB that Bill Gosper had to work with used to be a tremendous amount of memory, but it is the cache size of modern CPUs, which are typically paired with over 2,000 times that much RAM memory.

The reason hashlife uses so much memory is because there's no free lunch; if you ask it to generate 1,000 generations of a 1,000 by 1,000 highly random pattern, and the generation rule is such that the results are highly unpredictable, you will be consuming enough space to store a substantial fraction of all of those generations. Hashlife is amazing, but it is not magic. If you use a smaller step size, the memory consumption of hashlife is significantly mitigated (but so is its potential for galloping ahead.)

The fact that hashlife usually spends little time doing the primitive cell computation, and the fact that it essentially caches this computation across many uses, makes it possible to simulate significantly more complex rules effectively. This is the basis of Golly's extension to 256-state automata and the reason such perform so well in Golly. Again, highly random patterns can defeat this by causing generation of more leaves than fit in memory, but for most interesting patterns this does not happen. Golly uses smaller leaves for 256-state automata than it does for 2-state automata to help prevent this from occurring.

Normal life algorithms are generally highly parallel, so they are easily sped up as CPUs gain more cores and as highly parallel GPUs take over more and more the computation. So far, hashlife has been resistant to parallelization. A big challenge for the community is to write an effective, efficient, parallel hashlife, perhaps based on lock-free hashtables.

4 Comparisons

In this section we present a shootout between different algorithms for computing the game of life. This shootout is far from the last word, as we've selected only a handful of interesting patterns, and the algorithms themselves have different capabilities. We present this only to provide insight in algorithm performance.

4.1 Algorithm Implementations

The algorithms we compare are given here, in roughly increasing order of sophistication and performance. The first seven only support bounded universes; the last eight support unbounded universes (or

at least universes bounded only by the datatype used to represent coordinates). All implementations except the last four were written specifically for this paper with straightforward code; they might all benefit from various tuning or compiler optimizations, so the results should only be considered in the large. The last four algorithms are well-tuned implementations.

Bounded universe algorithms:

- `array4`: The naive schoolboy algorithm, implemented with static arrays and one byte per cell.
- `lookup`: The schoolboy algorithm but uses a single-cell lookup table and thus supports other outer totalistic rules.
- `lookup4`: Stores data as dense bitplanes, and uses a 2^{16}-element lookup table to compute the 2x2 result of a 4x4 square at once. This provides speed, while still permitting arbitrary rules to be used. This is most like what qlife and hlife use as their inner loop.
- `nybble`: Stores data as 16 cells per 64-bit long, and uses bit tricks to compute the next generation so is specific to GOL.
- `bitpar3`: Stores data as dense bitplanes, and does parallel bitplane addition. Specific to GOL.
- `sse3`: Similar to bitpar3 but uses SSE instructions to compute 128 bits at a time.
- `avx23`: Similar to bitpar3 but uses AVX2 instructions to compute 256 bits at a time.

Unbounded universe algorithms:

- `list3`: Simple list-based algorithm that lists the coordinates of cells that are alive. Supports unbounded universes. The actual computation is specific to the game of life but can be easily modified to be more general without much performance impact.
- `tree`: Uses 8x8 fat nodes with a tree container. Each fat node is computed with an algorithm like bitpar3.
- `list8x8`: Uses 8x8 fat nodes with a list container similar to list3. Fat node computation is the same as bitpar3.
- `list16x16`: Uses 16x16 fat nodes with a list container similar to list3. Fat node computation is similar to avx23.
- `qlife`: The non-hashing algorithm in Golly. Uses a tree structure (but not a quadtree. Each node splits space eight ways in one dimension). Performs period-two optimization. Computation is generic across algorithms without recompilation [2].
- `ulifelib`: Tree-based algorithm that uses AVX2 or SSE (depending on platform). Supports additional rules but only with recompilation. Performs period-two optimization [4].
- `hlife`: The hashing algorithm in Golly. Uses 8x8 leaves and supports general rules. We run it with 1GB RAM [2].
- `lifelib`: Hashlife algorithm that uses AVX2 or SSE (depending on platform). Supports additional rules but only with recompilation. We run it with 1GB RAM [4].

4.2 Patterns

- `r4kx4k`: A random 4000 by 4000 universe at 30% density. Stabilizes (with respect to a period of 2) after 12,000 generations (except for escaped gliders).
- `bp4kx4k`: A 4000 x 4000 universe filled with period-3 pulsars. Intended to evaluate the baseline performance of different algorithms that may have period-2 recognition. The hashlifes will run away quickly with this pattern.

	r4kx4k 2K	r4kx4k 16K	r4kx4k 4M	bp4kx4k 2K
array4	99.444	-	-	93.963
lookup	129.182	-	-	131.504
lookup4	10.683	85.827	-	10.706
nybble	7.513	60.151	-	7.443
bitpar3	2.377	19.073	-	2.379
sse3	1.195	9.456	-	1.238
avx23	0.782	6.372	-	0.827
list3	107.285	-	-	199.282
tree	10.945	99.114	-	10.488
list8x8	9.596	78.011	-	6.899
list16x16	2.196	18.486	-	2.070
qlife	3.804	5.696	140.768	9.205
ulifelib	2.491	3.429	82.889	6.842
hlife	53.533	68.816	68.943	0.083
lifelib	9.496	14.991	14.992	0.033

Table 1: Bounded vs. unbounded algorithms at 4K x 4K size. We include results at 4M for the random case to compare how the different sophisticated algorithms manage patterns that are primarily stable.

- `QGC1`: A difficult pattern that contains mostly closely-spaced glider streams that are going different directions; this gives hashlife a tough time.
- `breeder`: Gosper's original breeder. Fills one eighth of the universe with a wave of gliders. Population is quadratic in generations.
- `caterpillar`: A huge spaceship. The bounding box is about 4,195 x 330,721 (and thus requires 173MB as a single bitmap). There is a lot of regularity but the pattern is so large that the hashlifes don't easily run away.
- `jagged`: A pattern that grows linearly with generations and generates interesting curves as it runs.
- `lidka`: A small Methuselah. It expands rapidly kicking out many gliders and eventually settles down about generation 30,000.
- `mcc`: Metacatacryst, one of the early small quadratic-growth patterns.
- `spiral`: Similar to but smaller than QGC1.
- `unlim`: A pattern intended to produce unlimited novelty by positioning two rakes at right angles that emit gliders into the debris of each other.

4.3 Results

The first set of results we present is for large rectangular patterns that mostly stay rectangular, so we can compare the bounded and unbounded algorithms. For the bounded algorithms we allocate a universe just large enough to hold the pattern, and we ignore edge effects (mostly gliders that hit the edge and become blocks). The results are in Table 1.

With the exception of the four sophisticated algorithms, performance is very similar between the random initial state and the field of pulsars; this is because the simple algorithms are data-independent. They do not recognize pattern behavior. The first two algorithms that only compute a single cell at a time are very slow. The next five bounded algorithms are increasingly fast as they do more and more bits at once. The lookup4 algorithm, which does four output results at a time from a 4x4 input field,

	QGC1 128K	breeder 32K	caterpillar 1K	jagged 128K	lidka 1M	mcc 128K	spiral 512K	unlim 128K
list3	-	-	-	-	99.789	-	-	-
tree	231.231	263.220	147.303	166.304	38.217	130.305	238.966	557.311
list8x8	107.790	119.485	74.010	37.655	18.072	58.601	117.828	223.520
list16x16	65.802	43.734	25.808	29.624	6.245	26.492	67.860	87.257
qlife	112.341	170.113	95.010	76.559	3.474	10.779	104.610	22.847
ulifelib	62.401	141.140	41.081	79.610	1.860	5.822	60.214	13.066
hlife	57.239	0.018	25.654	0.100	0.320	0.215	68.581	2.974
lifelib	105.515	0.004	26.153	0.133	0.076	0.039	60.685	1.147

Table 2: Unbounded algorithm comparison.

but is easily generalized to other rules because of that lookup table, is almost as fast as the nybble algorithm but loses increasingly badly as we use 64-bit, 128-bit, and 256-bit logical operations.

For the unbounded algorithms, a fraction of the CPU time must be spent managing the information about what parts of the universe are alive and active (as well as handling the gliders that are shot off from the central section). For these dense patterns, tree and list8x8 are very similar, but list16x16 is much faster because of the AVX2 instructions it uses. The qlife algorithm, despite only doing four bits at a time in the leaves, eventually outperforms these simpler algorithms for the random case because it recognizes period 2 stability. Ulifelib beats qlife with its highly efficient leaf-level computations. (It is possible to run ulifelib in a way that it recognizes period-3 periodicity as well.)

Both hashlife algorithms are significantly slower than the sophisticated non-hashlife algorithms for the random universe while much random activity is still taking place. Lifelib soundly trounces hlife due to the much more efficient leaf-level computations. Note how the hashlife algorithms both take almost no time to advance from 16K to 4M generations, while the other algorithms either don't get there in our time limit or take more than a minute to do so. For the pulsar field, both hashlife algorithms easily and quickly run away.

Our second set of results are for more complex patterns. We chose patterns that are not trivial since most trivial patterns run too fast for meaningful comparison.

The simplest algorithm, list3, that uses individual cells rather than fat nodes, was not able to compete in any meaningful way with algorithms that used fat nodes. We include the lidka result just because the pattern had few enough live cells even at high generation counts where it was able to not lag too far behind.

For almost all patterns, the overall performance increased from the top of the table to the bottom as the algorithms increased in sophistication. For two of the patterns, QGC1 and spiral, the fast ulifelib algorithm was very close to the fastest hashlife algorithms (in this case hlife). The two sophisticated unhashed algorithms, ulifelib and qlife, were fairly comparable, although ulifelib generally won with its AVX2-powered routines specific to the game of life. There were a few patterns (breeder, caterpillar, jagged) where the simple list16x16 algorithm outperformed qlife and ulifelib. Where this happened, however, hashlife algorithms were usually even better.

For the hashing algorithms, lifelib beat hlife on breeder, lidka, mcc, and unlim; hlife won on QGC1, and they were close to tied on caterpillar, jagged, and spiral. Lifelib also makes more efficient use of memory, using 32-bit node indexes rather than 64-bit node pointers. On the other hand, hlife uses prefetching to overlap some memory references for hashtable lookups.

5 Discussion

It is clear from our results that there is no best Life algorithm; as you increase the sophistication of algorithms to take advantage of specific behaviors, you generally decrease the performance on the class of patterns that do not exhibit those behaviors. The attraction of hashlife is that it provides significant acceleration for a wide class of behaviors, even though for large random patterns it is

typically outperformed by simpler algorithms.

Equally clear is that the use of low-level bit tricks, and especially the wide bit operations available through SSE and AVX, can significantly improve performance. This remains true for hashlife algorithms; use of large leaves and efficient leaf calculations can help mitigate the overhead of creating nodes and canonicalizing them with a hashtable. Unfortunately the use of such bit tricks usually requires recompilation for different rules. On some platforms it may be possible to support just-in-time compilation of appropriate low-level code to provide a combination of performance and flexibility.

We might expect performance to increase even more as AVX is extended to 512 bits. Further increases are possible as newer instructions are introduced.

As GPUs take over from CPUs in terms of overall operations per second, effective use of GPUs for Life should be considered. Existing research focuses on computation on large bounded arrays without taking advantage of any periodicity or regularity. Can we exploit GPUs with more sophisticated algorithms and gain both of these advantages at once?

Even exploiting multiple cores in a CPU can be a challenge as the algorithms get more complex. In particular, no parallel versions of hashlife have been seen that outperform a single-threaded version.

These challenges will provide great entertainment and challenges for hackers in the decades to come, much as the generation of the algorithms described in this paper have provided so much fun for this author and many others in the past few decades.

References

[1] Fujita, Toru, Daigo Nishikori, Koji Nakano, and Yasuaki Ito, "Efficient GPU Implementations for the Conway's Game of Life", 2015 Third International Symposium on Computing and Networking.

[2] The Golly Team, Golly, `https://golly.sf.net`.

[3] Gosper, R. William, "Exploiting regularities in large cellular spaces", Physica D: Nonlinear Phenomena, v.10 pp. 75–80.

[4] Goucher, Adam P., LifeLib, `https://gitlab.com/apgoucher/lifelib`.

[5] Niemiec, Mark B., "Life Algorithms." Byte Magazine, January 1979.

[6] Rokicki, Tomas G., "An Algorithm for Compressing Space and Time", Dr. Dobbs, April 2006.

[7] Warren, Hank S. Jr., "Hacker's Delight", Addison-Wesley, 2012.

Numbers for Masochists: A Mental Factoring Cheat Sheet

by Richard Schroeppel and Hilarie Orman

Full paper at http://www.purplestreak.com/g4g13/mentalfactoring.pdf

Quick divisibility tests

2: low order digit of n is even. **3**: sum of the digits of n is divisible by 3.
5: low order digit is zero or 5. **7**: if $n = abc$, then n modulo 7 is $2 * a + bc$.
7, 11, and 13: $abcd$ modulo 1001 is $bcd - a$.
11: For abc, if $b = a + c$ or $b + 11 = a + c$, then $abc = 11 * ac$ or $11 * zc$ where $z = a - 1$. $11|abcd$ iff $a + c$ and $b + d$ are equal or if the difference is 11. **13**: $n/300 = \{q, r\}$, $13|(q + r)$ iff $13|n$.
37: If n has 3 digits, rotation preserves divisibility by 37.
97, 101, 103, ...: each $100 \pm n$ divides $10000 - n^2$.

Table 1: Useful and Memorable Multiples of Small Primes

	Column - high order digits, Row - units digit			
	1	3	7	9
1	1001	299, 1001	102, 1003, 6001, 10013	399, 1007, 1501, 7999, 10013
2		2001		2001
3	992, 3999, 10013		111, 999	
4	10004	301, 3999, 10019	10011	
5		1007, 10017		1003, 20001
6	10004		201	
7	994, 10011	511, 1022, 10001		1501, 3002
8		996, 20003		801
9			9991	
10	9999	9991	9951, 20009	981, 10028, 40003
11		1017, 20001		
12			1016, 8001	
13			10001	
24	964, 20003			

The Method for Factoring n: Using Table 2, select the quadratic form(s) and the term that is divisible by 5 (NB: if there is no entry for n, use the 120 Method and/or Difference of Squares). Solve each form modulo 100 using the fact that one of the squares is a multiple of 25. For each form, there will be one or two solutions < 25, call them r (and s). The candidates for the non-multiple-of-5 term are the set $\{50i \pm r, 50i \pm s\}$ such that the square is less than n (or $n/2$ or $n/3$).

For each candidate value, plug in its square into the quadratic form and solve for the square of the other variable. If that solution is, indeed, a square, and if $\gcd(x, y) = 1$, then the x and y values are a solution to the quadratic form.

If you find **two solutions**, the number is composite. Calculate the factors using vector addition/subtraction on the two solutions to minimize the result vector (u, v) and/or to have both terms divisible by 5. Divide both terms by $\gcd(u, v)$. Substitute u and v for x and y in the QF; the result will have a factor of n.

If all potential candidates less than the square root of n have been tried, and there is only **one solution**, then n is prime. If there are **no solutions**, n is composite; the factorization must be done with another method.

Example: 4469. Per table 2, we use the QF $x^2 + y^2$. Either $x^2 \equiv 0 \mod 100$ or $x^2 \equiv 25 \mod 100$. First assume 0 mod 100; then $r = 13$ because $13 * 13 \equiv 69 \mod 100$. The y candidates are $50j \pm 13$, and $y < 70$. Possibilities are 13, 37, and 63.

$4469 - 13^2 = 4300$ which is not a square.

$4469 - 37^2 = 4469 - 1369 = 3100$ which is not a square.

$4469 - 63^2 = 4469 - 3969 = 500$ which is not a square. Therefore, $x^2 \equiv 0 \mod 100$ is impossible.

Now assume $x^2 \equiv 25 \mod 100$; find r such that $r^2 \equiv 69 - 25 \mod 100 = 44$. That would be 12. The y candidates are $50j \pm 12$, $y < 70$: 12, 38, and 62.

$4469 - 12^2 = 4325$ which is not a square because the hundreds digit is odd.

$4469 - 38^2 = 4469 - 1444 = 3025 = 55^2$. This is a representation of 4469 as $55^2 + 38^2$.

$4469 - 62^2 = 4469 - 3844 = 625 = 25^2$.

Table 2: Properties of quadratic form terms

residue	low digit	quadratic form	5 divides	x parity	y parity	r^2 mod 100
1 mod 4	1 or 9	$n = x^2 + y^2$	either	either	1-p(x)	n, $n - 25$
1 mod 4	3 or 7	$2n = x^2 + y^2$	either	odd	odd	$n - 25$
3 mod 8	1 or 9	$n = x^2 + 2y^2$	y	odd	odd	$n - 50$
		$3n = x^2 + 2y^2$	x	odd	even	$(3n - 25)/2$
3 mod 8	3 or 7	$n = x^2 + 2y^2$	x	odd	odd	$(n - 25)/2$
		$3n = x^2 + 2y^2$	y	odd	even	$3n$
7 mod 24	1 or 9	$n = x^2 + 3y^2$	y	even	odd	$n - 75$
		$4n = x^2 + 3y^2$	y	odd	odd	$4n - 75$
7 mod 24	3 or 7	$n = x^2 + 3y^2$	x	even	odd	$n/3$
		$4n = x^2 + 3y^2$	x	odd	odd	$(4n - 25)/3$

Add the two representations $(55, 38)$ and $(25, 62)$ to get $(80, 100)$. The gcd is 20, dividing it out yields $(4, 5)$, $4^2 + 5^2 = 41$. By mental arithmetic, $4469/41 = 109$.

Filters. $n \equiv x^2 + y^2 \mod 3$. The squares modulo 3 are 0 and 1, the corresponding square roots are 0, ± 1. Let m be the residue of n modulo 3. List all solutions to $m \equiv u^2 + v^2 \mod 3$ using 0 and 1 for u^2 and v^2. When trying an x or y candidate, check that it is consistent with the solution set modulo 3. If it isn't, discard it. You can do the same thing modulo 9 (squares are 0, 1, 4, and 7), modulo 7 (squares are 0, 1, 2, and 4), or modulo 49 (squares are 0, $7j + \{1, 2, 4\}$).

Modulo 100 filters. Match the parity of the hundreds digits in n and the square of a candidate value. If y is an odd multiple of 5 and the QF is $x^2 + 2y^2$, use the pattern of thousands-hundreds digits. If the QF is $x^2 + 3y^2$ and the tens digit of n is odd, match the parity of the hundreds digit of $n - 25$ or $n - 75$ to the parity of the hundreds digit of the candidate.

Example: $1000009 = 1000^2 + 3^2$. From Table 2, $y^2 \mod 100$ is either 00 or 25.
$09 - 00 = 9 = x^2 \mod 100 \to r = 3$, and $09 - 25 = 84 = x^2 \mod 100 \to r = 22$, so the x candidates are $50 + 3$, $50 - 3$, $50 - 22$, $50 + 22$, $\ldots$; 50 ± 22 is modified to 100 ± 28 to match hundred's digit parity. Squares modulo 9 eliminate 997; squares modulo 7 and modulo 9 accept 972. $1000009 - 972^2 = 55225 = 235^2$. Combine $(1000, 3)$ with $(235, 972)$ to get factors 293 and 3413.

The 120 Method. Find solutions to $kn = ax^2 + by^2$ where k, a, and b are small. For each solution, add $-ab$ to the set Q and compute the closure of Q under multiplication, exact division, and division by a square.
For a $4i + 3$ number, if 2, 3, and 5 (irrespective of sign) are in Q, n can be factored or proved prime. For a $4i + 1$ number, if -1, 2, 3, and 5 are in Q, n can be factored or proved prime.
The trial divisors of n for a $4i + 3$ number: $120j + \{1, 49, d, e\}$ where $d = n \mod 120$, $e = 60 - 11d \mod 120$ and j goes from 0 to $\sqrt{n}/120$; for a $4i + 1$ number: $120j + \{1, 49\}$ where j goes from 0 to $\sqrt{n}/120$. Only prime divisors need be tested.

Example 2503: $n = 50^2 + 3 = 51^2 - 98 = 15 * 13^2 - 32$. The corresponding $-ab$ values are -3, 2, 30. By closure, $Q = \{2, 3, 30, 15, 5\}$. Then $d = 103$, $e = 7$; trial divisors are $120j + \{1, 7, 49, 103\}$. Testing 7 fails, 49 is composite, $103 > \sqrt{n}$. Therefore n is prime.

The Difference of Squares Method. Find x and y such that $n = x^2 - y^2$. One of the two squares will end in 00 or 25. Solve for the other square modulo 100 using the following equations.
For $n \equiv 1 \mod 4$:
$x \equiv 5 \mod 10$, $y^2 \equiv 25 - n \mod 100$
$y \equiv 0 \mod 10$, $x^2 \equiv n + 0 \mod 100$
For $n \equiv 3 \mod 4$:
$x \equiv 0 \mod 10$, $y^2 \equiv 0 - n \mod 100$
$y \equiv 5 \mod 10$, $x^2 \equiv n + 25 \mod 100$

Of the two solutions, one is based on x, the other on y. Use the solutions to build candidate sets of the form $\{50j \pm r\}$ as in The Method; one is for x candidates, the other is for y candidates. Alternate trying x candidates and y candidates, then change the limits for x and y as described next. If x and y both exceed their limits, then n is prime.

Limits for x and y. Use divisibility tricks to eliminate possible divisors up to $L = 37$. Call the upper limit for x L_x. $L_x = (L + n/L)/2$; the upper limit for y is $L_x - L$. To change the limits, use mental arithmetic to test more primes in sequence, set L to the last prime tested, and recompute the limits. Divisor restrictions (see full paper) can eliminate some primes without testing.

(Phi)ve is a Magic Number

James J. Solberg
Professor Emeritus, Purdue University
Written for the 13th Gathering for Gardner, 2018

Martin Gardner would surely have noted the centennial significance of January 6 of this year. On that day, the digits of the date 1/6/18 matched 1.618 – the leading digits of the famous golden ratio number Phi. If you wanted three more digits, you could have set your alarm to celebrate at 3:39 am or at a more reasonable 3:39 pm when the eight digits of that moment aligned to 1.6180339.

I am sure that Martin would have also noticed the puns in my title. The first word refers to both Phi and five, and the fourth word refers to magic squares. My goal is to reveal several surprising connections between the two values and their powers using both Fibonacci numbers and five-by-five magic squares. A warning: if ordinary word puns tend to make you numb, then my mathematical puns will make you number.

As you presumably know, the golden ratio number Phi (or ϕ) shows up in nature, in art, in architecture, and even beauty salons, as well as mathematics. If you are not familiar with ϕ, you are in for a treat. There are plenty of books and websites that explain the fascinating properties and ubiquitous nature of this mysterious number. I want to use a few of the many identities, so here is a short summary.

One of the ways to determine the golden ratio is to find the place to divide a line so that the ratio of the length of the longer segment to the shorter is the same as the ratio of the whole line to the longer segment. If x is the length of the longer segment and y is the length of the shorter, equating those two ratios gives the equation $\frac{y}{x} = \frac{y+x}{y}$ or $y^2 = xy + x^2$. We are only interested in the ratio, so we can arbitrarily set the value of the smaller value x to 1, to get a quadratic equation in a single variable, $y^2 = y + 1$, or $y^2 - y - 1 = 0$. Then you can find the two roots using the familiar quadratic formula, $\frac{-b \pm \sqrt{b^2-4ac}}{2}$. There is one positive root, namely, $\frac{1+\sqrt{5}}{2} = 1.6180339...$, which is the golden ratio, ϕ. The other root is $\frac{1-\sqrt{5}}{2} =$ -0.6180339..., which can also be expressed as $1 - \phi$ or also as $\left(-\frac{1}{\phi}\right)$ or $(-\phi)^{-1}$.

The identities that I will use later are $\phi^2 = \phi + 1$ and $\phi^{-1} = \phi - 1$. Both of these are directly obtainable from the above. Also, instead of expressing ϕ as a fraction and a square root, we could use decimals and powers expressed as decimals, giving the interesting form:

$$\phi = 0.5 + 0.5(5)^{0.5}$$

It is not the easiest form to remember, but it does display the intimate connection between ϕ and 5. This is only the first of many links to be discovered.

Phi, Five, and Fibonacci

As you probably know, Fibonacci numbers are a sequence of integers defined by the rule that each number in the sequence is the sum of the previous two. If F_n denotes the n^th^ Fibonacci number, $F_n = F_{n-1} + F_{n-2}$. Along with that rule, you must know that the first two Fibonacci numbers are 1's, i.e., $F_1 = 1$ and $F_2 = 1$. So the sequence begins {1, 1, 2, 3, 5, 8, 13, ...}.

If you are interested in how rapidly that sequence is increasing, you could examine the ratio of each Fibonacci number to the previous one, i.e., $\frac{F_n}{F_{n-1}}$. For example, $\frac{F_2}{F_1} = 1$, $\frac{F_3}{F_2} = 2$, $\frac{F_4}{F_3} = 1.5$, and so forth. Most people who are familiar with both ϕ and the Fibonacci numbers are aware the limiting value of those ratios is ϕ. That is, $\lim_{n\to\infty} \frac{F_n}{F_{n-1}} = \phi$. Fewer people seem to be aware that there is a formula that expresses all of the Fibonacci numbers *exactly* in terms of ϕ. It is known as Binet's formula, published in 1843, although it was actually found and published much earlier by both Euler in 1756 and de Moivre in 1730.[1] Here it is:

$$F_n = \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}}$$

This is an interesting result because the square root of 5 and all the non-zero powers of ϕ are irrational, yet the parts combine to produce exact integers for every value of n. Although we normally think of the Fibonacci sequence as containing only positive integers, the same formula works for zero and negative n, so the sequence can extend 'backwards.' That is, $F_0 = 0$, $F_{-1} = -1$, $F_{-2} = -1$, $F_{-3} = -2$, and so forth. As you can see, the negative sequence just mirrors the ordinary positive one.

Putting this formula into a form that emphasizes ϕ's and 5's, the result can be expressed as:

$$F_n = 5^{-0.5}(\phi^n - (-\phi)^{-n})$$

The 5^th^ Fibonacci number happens to be 5, so 5 and ϕ are linked by this surprising equation.

$$5 = 5^{-0.5}(\phi^5 - (-\phi)^{-5})$$

By the way, the seventh Fibonacci number is 13, so in honor of the thirteenth gathering,

$$13 = 5^{-0.5}(\phi^7 - (-\phi)^{-7})$$

Generalized Fibonacci sequences use the same recursion rule (the sum of the previous two) but with arbitrary starting values. If the n^th^ number in the generalized Fibonacci sequence is G_n, and the starting values are $G_0 = x$ and $G_1 = y$ for any x and y, then

[1] Deriving this formula using standard methods for solving a second order difference equation with two initial conditions is an easy and very satisfying exercise, suitable for high school algebra students.

$$G_2 = x + y$$
$$G_3 = x + 2y$$
$$G_4 = 2x + 3y$$
$$G_5 = 3x + 5y$$

and so on. You can see from the way the coefficients of x and y are developing that the generalized Fibonacci numbers are closely related to the ordinary Fibonacci numbers by the equation $G_n = F_{n-1}x + F_n y$. The formula works for negative n and even for non-integer x and y. All of these results are well known, along with many other interesting features of the Fibonacci numbers.

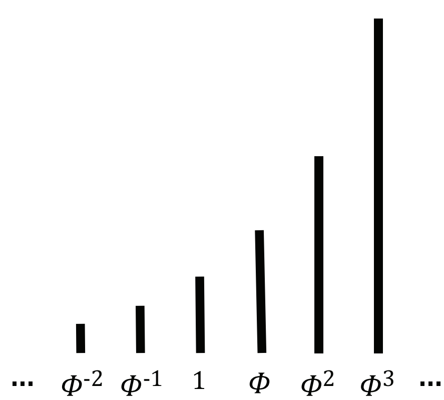
One particular instance of the generalized Fibonacci sequence that is relevant in the present context uses the initial values 1 and ϕ. I will call that sequence of numbers the Phi-bonacci sequence[2] and designate the terms as ϕ_n. So,

$$\phi_0 = 1$$
$$\phi_1 = \phi$$
$$\phi_2 = 1 + \phi$$
$$\phi_3 = 1 + 2\phi$$
$$\phi_4 = 2 + 3\phi$$

and so forth. In general, for all n, $\phi_n = F_{n-2} + F_{n-1}\phi$. The reason that sequence is interesting and relevant is that the closed form solution to the difference equation is very simple. It is:

$$\phi_n = \phi^n$$

That is, the n^th number in the sequence is the n^th power of ϕ. The same formula works for negative n, so $\phi_{-n} = \phi^{-n}$. When you think of ϕ as the golden ratio representing the ideal proportioning of line segments, the Phi-bonacci sequence is a set of line segments extending to infinity in both directions in which each one is multiplied by ϕ to match the larger neighbor and divided by ϕ to match the smaller neighbor.



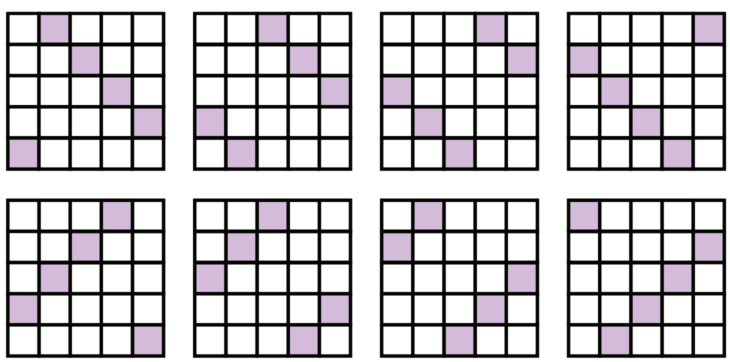
[2] Others have used this term for different sequences, but I think the use in this context is clear.

Phi, Five, and Magic Squares

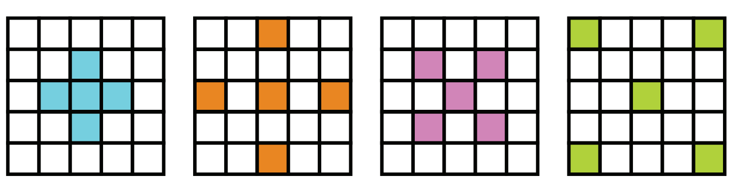
The following is a magic square of order five whose magic sum is 1.6180. Every row, column, and diagonal adds to the first five significant digits of ϕ.

0.3224	0.3246	0.3243	0.3235	0.3232
0.3238	0.3230	0.3227	0.3244	0.3241
0.3247	0.3239	0.3236	0.3233	0.3225
0.3231	0.3228	0.3245	0.3242	0.3234
0.3240	0.3237	0.3229	0.3226	0.3248

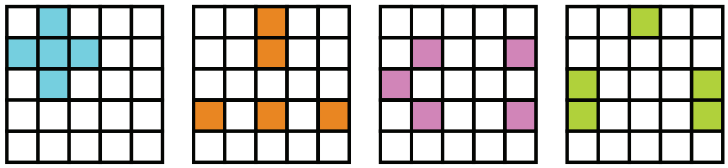
However, it is not just an ordinary magic square. It consists of the twenty-five consecutive decimal numbers starting with 0.3224 in the upper left cell and ending with 0.3248 in the lower right cell. It is pandiagonal, which means that all eight of the broken diagonals also sum to the same value.



You can also shift the rows or columns in any direction, wrapping around as you do so, and the square will remain magic. In addition to the twenty 'magic' patterns that you get from the five rows, five columns, five left diagonals, and five right diagonals, there are four that involve the center cell and four symmetrically surrounding cells.

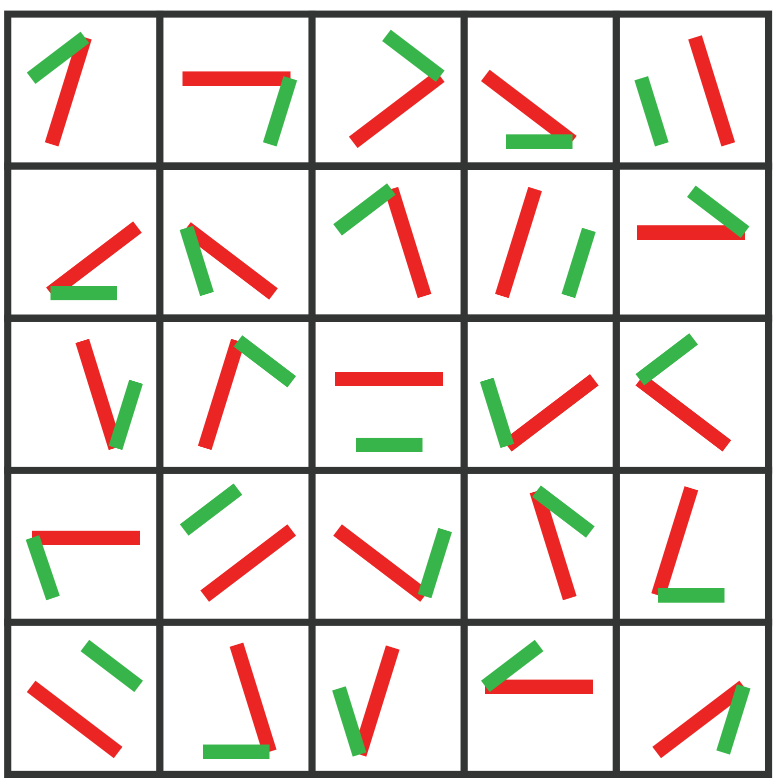


Each of these can be shifted so that center is in any of the 25 cells (wrapping around as necessary) so each of these four patterns has 25 variations, for a total of 120 patterns whose cell values add to the magic sum of 1.6180. For example, the values in the following patterns match the magic sum.

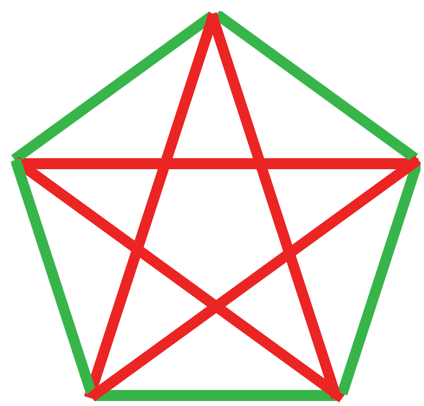


Furthermore, the magic square is symmetric, which means that any two cells that are at an equal distance from the center on a straight line through the center will contain values that sum to 40% of the magic sum, or 0.6472. That implies that any such opposite pair, together with any other such opposite pair and the center cell will match the magic sum. There are actually 780 combinations of five cells that match the magic sum! (You may use a calculator or a spreadsheet to check if you wish.)

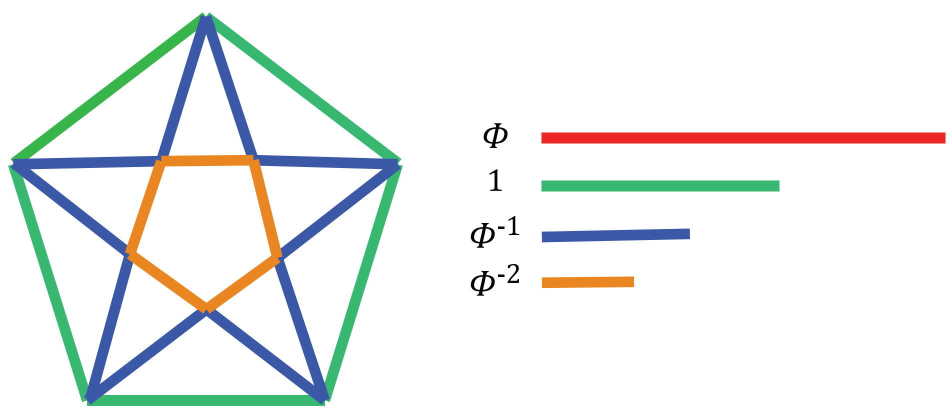
Of course, a purist would point out that this square does not really involve the true ϕ, but only a truncated approximation of it. In order to correct that flaw, I will resort to geometry. Each cell will contain two line segments: a red one of length ϕ and a green one of length 1 (in arbitrary units). When we add the geometric entries in cells, we simply superimpose them. With that geometric definition of addition, the following square is magic.



The square is pandiagonal, so all of the broken diagonals combine to the same result. The four patterns around the central cell, as well as the shifted versions also produce the pattern, so you get 120 ways that five cells form the magic sum! All 120 patterns sum to same figure, a five-pointed star (or pentagram) inscribed in a five-sided regular polygon (or pentagon). This symbol has been widely used throughout history to represent magical powers and as an instrument for casting magic spells, either evil or to protect against evil. More germane to our interest here, it contains many instances of ϕ and its powers.



We already know that every red line is of length ϕ. If each red line is broken into the three segments at the crossing points, every blue line segment is of length ϕ^{-1} and every orange segment is of length ϕ^{-2}.



Alternatively, if you want to measure all of the line segments relative to the shortest orange ones, taking those as the unit of measure, the blue lines would have length ϕ, the green ones would have length ϕ^2, and the longest red lines would have length ϕ^3. All of these powers of ϕ are successive numbers in the Phi-bonacci sequence.

The magic square is pandiagonal, so all of the broken diagonals combine to the same result. The four patterns around the central cell, as well as the shifted versions also produce the pattern, so you get 120 ways that five cells form the pentagram-in-pentagon figure.

If you take the numerical values of the line segments in the magic sum, they total $5 + 5\phi$, or using the identity $\phi^2 = \phi + 1$, the magic sum is $5\phi^2$. If you calculate what that amounts to, it is slightly more than 13, which seems fitting for the thirteenth G4G. If you add up the values of all of the line segments in the entire square, the total is $(5\phi)^2$, which seems appropriate for a five-by-five magic square designed to celebrate 5 and ϕ. It seems only fair to leave some fun for others, so I will only suggest that the perimeters and areas of the many internal polugons contain powers of ϕ and 5.

Finally, I would like to point out that Martin Gardner published his column in Scientific American starting in 1956 and ending in 1981—a total of 5^2 years. Can there be any doubt that (Phi)ve is a magic number?

References:

There are many books and articles about Phi, Fibonacci numbers, and magic squares. A quick search on the internet will produce enough to keep you entertained for weeks. If you want to learn how I created the magic squares (easily), the method is fully explained in my 2017 book, More Magic Square Methods and Tricks,

List of Authors | VOLUME 1

List of Authors (cont.)

G4G13 Exchange Book

Volume 1

Art, Games, Magic, & Math

Atlanta, Georgia

APRIL 11 - APRIL 15, 2018

Welcome to this glimpse of some of the fun and excitement of the 13th Gathering for Gardner (G4G13) in Atlanta, Georgia, April 11-15, 2018. Here you will find the program of events, and 78 papers that are write-ups by many of the presenters who made this event so vibrant. The subjects are far-ranging, all touching on subjects that fascinated Martin Gardner. Placed into sections on Art, Games, Math, and Magic... these papers describe puzzles, games, illusions, magic, and curiosities both mathematical and otherwise.

- excerpt from the Preface by Doris Schattschneider

Published by: Gathering 4 Gardner™