PI IN THE SKY

Issue 13, Fall 2009

In This Issue:

Pacific Institute *for the* Mathematical Sciences

On the Cover

The Egyptian game Senet, one of the oldest board games, dates back as early as 3500 BCE. This board, from the Brooklyn Museum, is inscribed with the name, Amunhotep III.

Contact Information

Pi in the Sky
PIMS University of Victoria Site Office, SSM Building Room A418b
PO Box 3060 STN CSC, 3800 Finnerty Road, Victoria, BC, V8W 3R4
T/(250) 472-4271 F/(250) 721-8958, E-mail: pims@math.uvic.ca

Significant funding for *Pi in the Sky* is provided by Alberta INNOVATION AND SCIENCE

Pi in the Sky is a publication of the **Pacific Institute for the Mathematical Sciences** (PIMS). PIMS is supported by the Natural Sciences and Engineering Research Council of Canada, the Province of Alberta, the Province of British Columbia, the Province of Saskatchewan, Simon Fraser University, the University of Alberta, the University of British Columbia, the University of Calgary, the University of Lethbridge, the University of Regina, the University of Victoria and the University of Washington.

Pi in the Sky is aimed primarily at high school students and teachers, with the main goal of providing a cultural context/landscape for mathematics. It has a natural extension to junior high school students and undergraduates, and articles may also put curriculum topics in a different perspective.

Submission Information

For details on submitting articles for our next edition of *Pi in the Sky*, please see:

http://www.pims.math.ca/resources/publications/pi-sky

Editorial: *David Leeming, Managing Editor, Pi in the Sky*

We welcome our readers to another issue of Pi in the Sky. In addition to being on-line at www.pims.math.ca/pi , we have about 1200 subscribers in 56 countries.

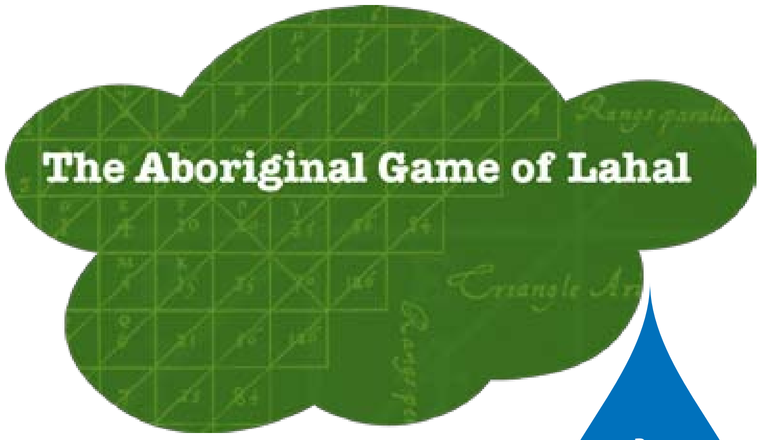
Once again, we were faced with the challenge of trying to publish articles that can be understood by capable high school and College-level students. In this context, we are constrained by the articles we receive from authors. This issue contains articles on mathematical games, geometry, face recognition, the Laws of Physics and elevator rides (in purgatory!). We also have a book review written by high school students.

In this issue, we announce the winners of the Math Challenge posed in Issue #11 (Spring 2008).

One of our Editors, Volker Runde, has resigned. Volker has served the magazine well for a number of years and we thank him for his dedicated service to Pi in the Sky. We welcome Murray Bremner, University of Saskatchewan, to the Editorial Board.

This is my last issue as Managing Editor of Pi in the Sky. I have thoroughly enjoyed my five years in the position. I would like to acknowledge the support of my Editorial Board for reviewing articles submitted and for their efforts in proofreading drafts of pending issues. I leave the magazine in the capable hands of Anthony Quas, University of Victoria, who will become Managing Editor of Pi in the Sky starting with Issue #14.

Table of Contents



The Aboriginal Game of Lahal

By
Peter J. Dukes
and
David J. Leeming[1]

A Simple Version of Lahal

In [1], the authors state that "Mathematics is a subject where Aboriginal students can feel particularly isolated and alienated." There are, however, some examples of Aboriginal games and activities that provide opportunities for exploration of the underlying mathematical ideas. One such game is Lahal, also known as 'the bones game.' It is usually played by two teams and each team could have anywhere from three to twenty members. The game consists of bones (typically two male, two female) and sticks (six or more). An excellent description of this and other versions of Lahal is contained in [1].

Here, we will present a simple version of Lahal for two players. Each player begins with six sticks and one bone. They takes turns with one player (the hider) hiding the bone in one of his hands and the other player (the pointer) guessing which hand contains the bone. If the pointer guesses correctly he gets a stick from his opponent. If he guesses incorrectly, the hider gets a stick from his opponent. The game continues until one of the players has all twelve sticks. Naturally, we assume a 50% probability that the pointer can correctly guess the hand containing the bone.

The outcomes of the first six rounds of this game are shown in Table 1. Specifically, the entry in the k^{th} column of round n indicates the number of possible game sequences which arrive at the outcome of one player holding k sticks. It is easy to compute each successive row in the table as it is the sum of the entries to the right and left in the row above. So, for instance, we have, in round 5 (row five - we count the first row as row zero) the entry 10 in column k=5. This is the sum of the numbers 4 and 6 above it in row 4. This corresponds to ten (5, 7) stick arrangements, coming from four (4, 8) arrangements and six (6, 6) arrangements in the previous round.

The array of numbers in Table 1 is known as Pascal's Triangle. Later, we will describe another way to find the entries in Pascal's Triangle using binomial coefficients. Note that the sum of all the entries in each row is a power of 2, more precisely, in row n the sum of the entries is 2^n. To calculate the probability of holding k sticks after n rounds, $n \leq 6$, we divide the entry in the k^{th} position by 2^n.

For example, after six rounds, since there are 2^6=64 different ways to arrive at one of the seven outcomes, so the probability of winning all six sticks from the pointer (or losing all six sticks to the pointer) is 1/64 and the probability that each player still holds six sticks is $20/64 = 5/16$.

Note that it takes an even number of turns to win a game. (Why?) After six rounds, the hider could be holding 0, 2, 4, 6, 8, 10 or 12 sticks. After six rounds, the game may be over, with the hider winning or losing. In case the game is not over, we cannot rely entirely on Pascal's Triangle any longer to calculate the probabilities of a win or a loss. Let us assume now that the game proceeds to eight or more rounds. The number of ways of reaching one of the outcomes after seven, eight, nine or ten rounds is shown in Table 2 below the line.

To calculate the probability of a particular outcome, for example, that each player has six sticks after eight rounds, we find the sum of the numbers in row eight which will be the denominator of the fraction we use. That sum is 248, so the probability that each player has six sticks is $70/248 = 0.282$. Using the same reasoning (you can fill in the steps)

http://brownsharpie.courtneygibbons.org/

[1]Both authors are with the Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., CANADA V8W 3R4
e-mail: dukes@uvic.ca, leemingd@uvic.ca

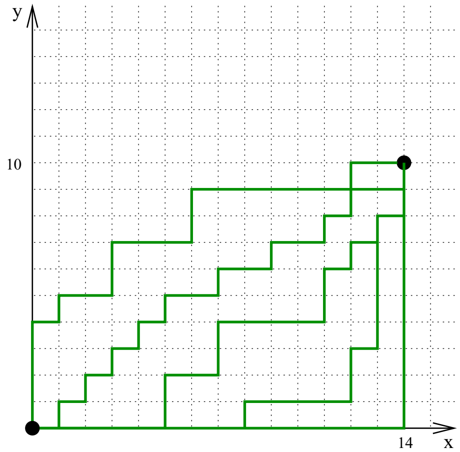


Figure 1: Various minimal lattice paths

we see that after ten rounds, the probability that each player has six sticks is $252/944 = 0.267$. Calculate the probability that each player has six sticks after twelve rounds, or after fourteen rounds. What is happening to these probabilities as the number of rounds increases?

We could continue this process to find the probabilities for the number of sticks held by the hider and the pointer after any even number of rounds. However, there are some obvious questions that any mathematician reading this would be asking by now.

1. Is there a formula for computing the number of sequences resulting in a given number of sticks for each player?

2. Will the game almost surely end in a finite number of turns?

3. If not, is there a limiting value to the probabilities?

The answers to these questions requires probing deeper into the mathematics behind this simple game. We will pursue these and other questions in Section 2 of this article.

If you are familiar with binomial coefficients, then you may wish to go directly to the next Section. If you are not, then this brief introduction may be useful to you in reading what follows.

First, let us examine factorials. For any non-negative integer n, the number n factorial whose symbol is $n!$ is defined by $0! = 1$, $1! = 1$, $n! = n(n - 1) \dots (2)(1)$, so for instance $4! = 24$, $9! = 362880$, and so on.

For nonnegative integers n and k, $0 \le k \le n$, the ***binomial coefficients***, denoted by $\binom{n}{k}$ (and read 'n choose k') give us the $n + 1$ entries in the n^{th} row of Pascal's Triangle. A very useful identity in calculating the binomial coefficients is $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$. See if you can prove this identity using the factorial form $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. From this, we see that $\binom{n}{k}$ actually counts the number of k-element subsets of an n-element set. For this reason, there is a standard convention that $\binom{n}{k}$ is deemed equal to zero if $k > n$ or $k < 0$. Pascal's Triangle gives us a triangular display of the nonzero binomial coefficients (see Table 1).

Analysis and Generalizations

It is helpful to draw a diagram which depicts a game such as Lahal. Call the two players X and Y, and associate them with the x- and y-axes of a Cartesian grid. Suppose instead of sticks trading back and forth that we simply keep track of how many times Y has won, versus how many times X has won. Each point (a, b) in the Cartesian plane, where a and b are nonnegative integers, corresponds to a wins for player X and b wins for player Y. There have been $a+b$ rounds played so far, and each player has either gained or lost $|a\text{-}b|$ sticks. In this model, the 'equilibrium' state is the line $y=x$, where each player has the same number of sticks.

Each game of Lahal unfolds as a 'path' in the Cartesian grid, from $(0, 0)$ to (a, b), the current situation. Moves along these paths are either east or north, according to whether X or

	0	1	2	3	4	5	6	7	8	9	10	11	12
Start							1						
Round 1						1		1					
Round 2					1		2		1				
Round 3				1		3		3		1			
Round 4			1		4		6		4		1		
Round 5		1		5		10		10		5		1	
Round 6	1		6		15		20		15		6		1

Table 1: Pascal's Triangle: sticks versus rounds, 0 to 6

	1	2	3	4	5	6	7	8	9	10	11
Round 5	1		5		10		10		5		1
Round 6		6		15		20		15		6	
Round 7	6		21		35		35		21		6
Round 8		27		56		70		56		27	
Round 9	27		83		126		126		83		27
Round 10		110		209		252		209		110	

Table 2: sticks versus rounds, 5 to 10

$$\sum_{i=-\infty}^{\infty} \binom{2n}{n+12i} + \binom{2n}{n+12i+1} + \binom{2n}{n+12i+2} - \binom{2n}{n+12i+6} - \binom{2n}{n+12i+7} - \binom{2n}{n+12i+8}$$

Figure 2: $f(n)$, the total number of non-ending sequences after $2n$ rounds rounds

Y (respectively) wins a given round. Such paths, which never 'backtrack' are called ***minimal paths***. Just as in our early discussion, there are $\binom{a+b}{a}$ paths from (0, 0) to (a, b). This is a^{th} entry in the $(a+b)^{\text{th}}$ row of Pascal's Triangle, where indexing entries and rows begins with 0. Figure 1 shows various minimal paths from (0, 0) to (14, 10).

One player wins when the game's path reaches a 'boundary,' which is a certain diagonal line. We must now discuss these boundaries further.

A minimal lattice path is ***subdiagonal*** if it lies on or below the line $y = x$. (In Lahal, this means one player is always winning throughout the game.) Counting the number of subdiagonal minimal lattice paths is a famous problem in combinatorics. It turns out that the number of paths on or below line $y = x$, from (0, 0) to (n, n), equals the n^{th} Catalan number, or

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

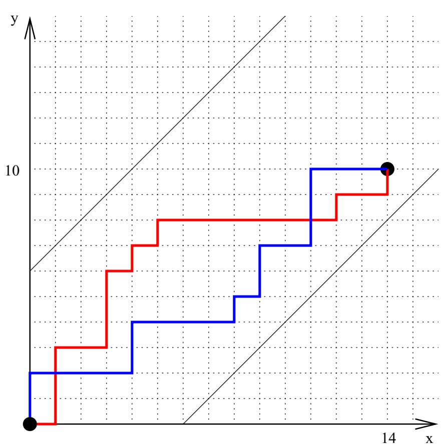


Figure 3: Two minimal lattice paths within boundaries

See the excellent reference [2] for more details on minimal lattice paths and the Catalan numbers.

Let's return to Lahal. Suppose in a general Lahal game with k sticks per player, that after $2n$ turns one player has $k + r$ sticks, and the other has k - r sticks, $0 < r < k$.

The number of ways this can occur is in one-to-one correspondence with the number of minimal lattice paths from (0, 0) to $(n + r, n - r)$ which do not intersect either of the lines $y = x - k$ and $y = x + k$. Like the Catalan restriction above, this is another problem on lattice paths, now with two linear boundaries. See Figure 3 for $n = 12$, $r = 2$, $k = 6$.

By results in [3] the number of such paths is

$$\sum_{i=-\infty}^{\infty} \binom{2n}{n+r+2ik} - \binom{2n}{n+r+(2i+1)k}$$

Note that the sum is finite, since almost all terms are zero.

In our case, we apply this formula for k=6, and r=0, 1, . . . , 5 to obtain the distribution for the $2n$th round. For instance, the number of game sequences which result in two players with 6 sticks each (i.e. $r = 0$) after $2n = 20$ rounds is

$$\binom{20}{10} - \binom{20}{4} - \binom{20}{16} = 175\,066.$$

The behavior after an odd number $2n + 1$ of rounds can be determined from round $2n$ using the 'sum rule' which generates Tables 1 and 2.

The case $r = 6$ ends the game, and one cannot explicitly apply the formula in this case.

There are 2^{2n} possible sequences for $2n$ coin flips (or wins/losses by the Lahal pointer). Some number $f(n)$ of these sequences result in the game *not* being over by the $2n^{\text{th}}$ round. This $f(n)$ can be found using the formula above, and is shown fully in Figure 2. (Who would have thought that such a simple game leads to such a long formula!?) The probability of the game ending on or before the $2n^{\text{th}}$ round is

$$Pr = 1 - 2^{-2n} f(n).$$

This actually tends to 1 as $n \to \infty$. Therefore, a game of Lahal (almost surely) will end.

This is rather intuitive. If a fair coin is tossed repeatedly, you should expect *eventually* a run of 12 (or even a million) heads in a row. This outcome delivers the win for one player over the other, no matter how many sticks are in play!

References

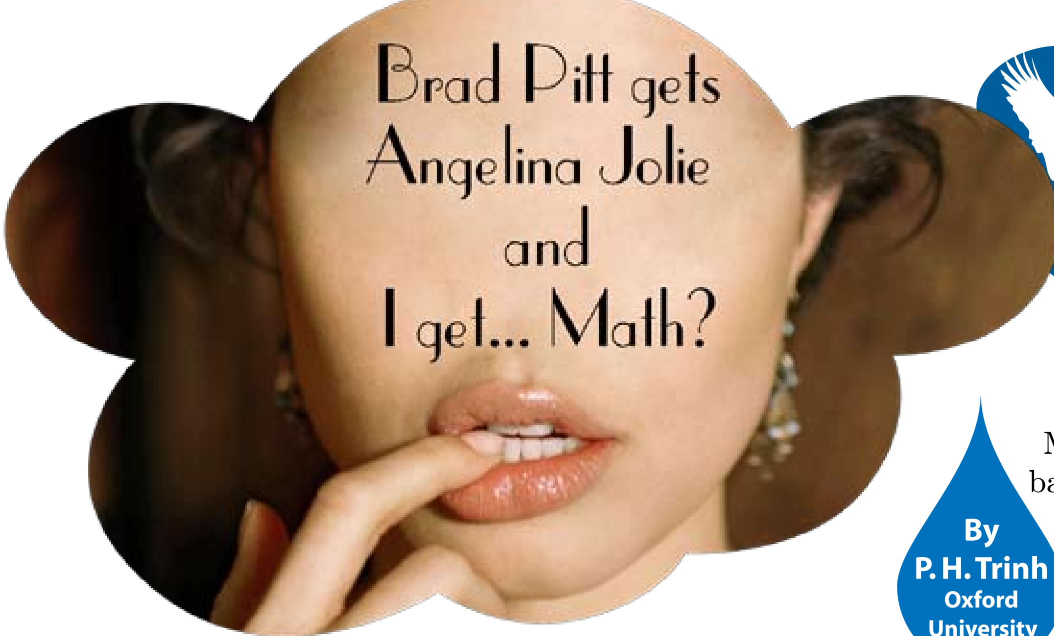
[1] T. Pelton, L. Francis-Pelton and L. Williams, Exploring probability with aboriginal games, *Vector*, BCAMT, *Spring 2006*.

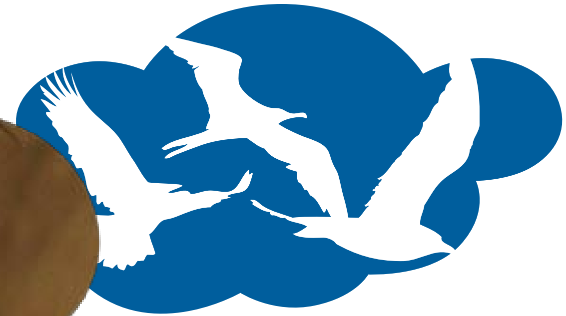
[2] R. Brualdi, *Introductory Combinatorics*, 4th ed. Prentice Hall, New Jersey, 2004.

[3] S. Dua, S. Khadilkar and K. Sen, A modified random walk in the presence of partially reflecting barriers. *J. Appl. Prob. 13 (1976)*, 169-175.

[4] Pascal's Triangle on Wikipedia: http://en.wikipedia.org/wiki/Pascal's triangle

Brad Pitt gets Angelina Jolie and I get... Math?





Who in the Land is Fairest of All?

By
P. H. Trinh
Oxford University

Seventh Grade Blues

Back in the seventh grade, one of the girls told me I looked like Keanu Reeves. Seriously, I was hanging upside-down on the jungle gym, minding my own business, and she just walked over and told me like it was no biggie.

Then she giggled like a moron and ran away.

Now while this singular moment of brazen flattery would become the highlight of my paste-eating academic career, I was also torn. On one hand, I wondered how *anyone* could confuse Keanu (black shades, gothic trench coat, totally awesome) with me (pubescent, angst-ridden, gawky)? Was this all some awfully cruel and sadistic joke girls liked to play on unsuspecting boys?

How might we go about identifying that person from a large database of faces?

On the other hand, maybe — *maybe* she was on to something. Maybe somewhere — *somehow*, behind all that bad acne and ruffled hair, my hidden Keanu-like features beckoned faintly, like some distant lighthouse obscured by fog.

Today, however, thanks to the latest advances in facial recognition, I no longer have to wonder: she was right.

A more computationally efficient way would be to examine these faces as a statistical whole rather than as the sum of their parts.

MyHeritage.com is an internet-based company that offers you the chance to see which celebrity you most resemble. Remember how in Snow White, the queen has a magical mirror which provides her with uninhibited flattery? This is the same, but like, tons better.

After a free signup, you upload a large-ish jpeg of your mug, then let the software crank away. My personal resemblance results were: Brad Pitt (71%), Keanu Reeves (63%), Luke Perry (63%), and Matt Damon (63%).

Brad Pitt? *Really*? Matt Damon? *Really*? Who *wouldathunk*? But y'know, as I gaze into the mirror... well... yes, I see it now. Definitely. We're practically brothers!

How does it all work? Is this actual science or just deceptive flattery? To understand how facial recognition works, we're going to have to delve into the mathematics behind the algorithm.

Recognizing Faces

Suppose we were given someone's picture. How might we go about identifying that person from a large database of faces?

One way we can go about it is by identifying the characteristics of the subject

One way we can go about it is by identifying the characteristics of the subject — perhaps the person has small lips, or a pointed chin, or distinctive eyes. From here, we then consult the database, going from picture to picture, each time isolating the features of the faces and checking for a match.

Whilst this might work, it would also be a lot of work; algorithms would need to be defined to analyse each desired feature. Imagine having to do this for each of the thousands of faces that stream pass the gates at a football match or in a busy airport.

As stored in a computer, a picture is nothing more than a great big grid of dots (or pixels).

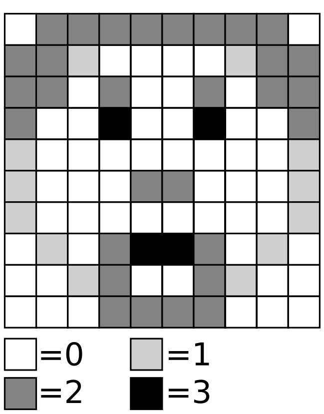
Now in the abstract theory of linear algebra, these grids of pixels are called 'vectors'.

A more computationally efficient way would be to examine these faces as a statistical *whole* rather than as the sum of their parts. This is similar to the difference between identifying a city by its landmarks and identifying the same city by the density of its roads, the clusters and heights of its buildings, its downtown areas, and so on.

A Picture is Worth a Thousand Digits

Snap! But what are pictures, *really?*

As stored in a computer, a picture is nothing more than a great big grid of dots (or pixels). If the picture is greyscale, each pixel is associated with a number from 0 to 255 representing its brightness, from pitch black (0) to pure white (255).



0	2	2	2	2	2	2	2	2	0
2	2	1	0	0	0	0	1	2	2
2	2	0	2	0	0	2	0	2	2
2	0	0	3	0	0	3	0	0	2
1	0	0	0	0	0	0	0	0	1
1	0	0	0	2	2	0	0	0	1
1	0	0	0	0	0	0	0	0	1
0	1	0	2	3	3	2	0	1	0
0	0	1	2	0	0	2	1	0	0
0	0	0	2	2	2	2	0	0	0

Figure 1: A picture is nothing more than a large grid of numbers.

Now in the abstract theory of ***linear algebra***, these grids of pixels are called ***vectors***. You've probably encountered vectors before in Physics class and in fact, these 'face vectors' are quite similar.

Like vectors representing force or motion, these new 'face vectors' have a ***magnitude*** (an overall brightness), as well as a ***direction***. Moreover, they can be added, subtracted, multiplied, and manipulated like most other mathematical quantities — the only difference is that they inhabit some higher-dimensional ***face space***, rather than the two or three dimensional physical world we live in.

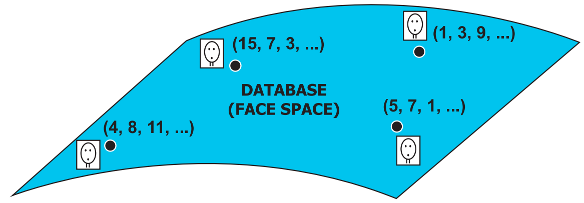


Figure 2: Faces can be identified as coordinates in a higher dimensional plane.

What's Your Eigenface Basis?

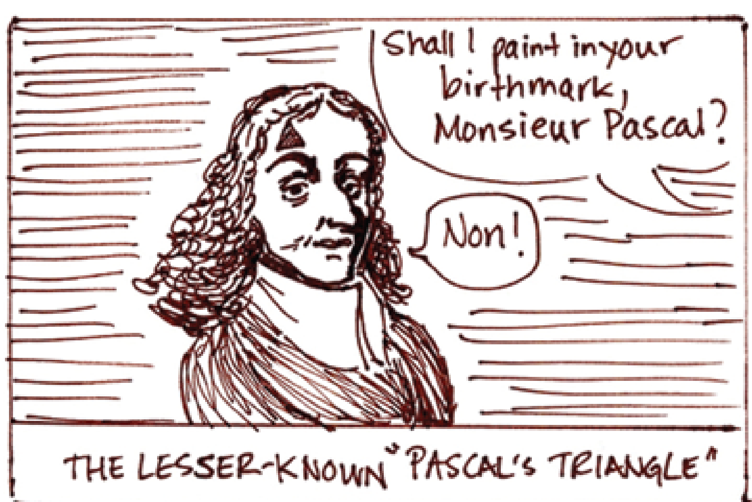
However, face spaces are complicated affairs — they're high dimensional boxes stuffed with a large number of faces, each face containing thousands of pixels.

...we look to construct a small group of pictures representing the general facial patterns of the database. This small but crucial group is called the eigenface basis.

It would thus be foolish to try and compare each face pixel by pixel; instead we look to construct a small group of pictures representing the general facial patterns of the database. This small but crucial group is called the ***eigenface basis***.

Think of how, when we analyse the motion of a ball flying through the air, we break the motion into its horizontal and vertical components. These two components provide a fundamental basis capable of describing any arbitrary motion.

Similarly, once the eigenface basis is found using linear algebra, each face in the database can then be expressed using certain percentages of each of



http://brownsharpie.courtneygibbons.org/

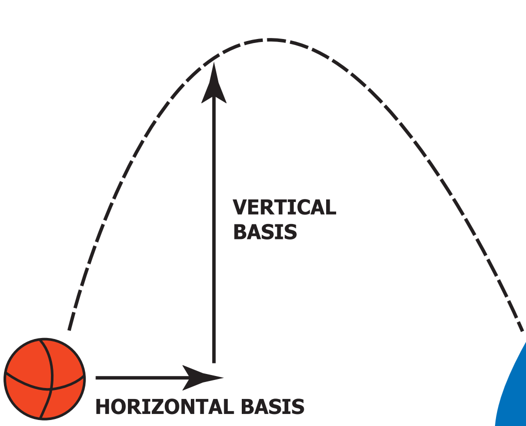


Figure 3: Vectors can be decomposed into basis elements.

the building blocks. For example, we may say that a picture is composed of 10% of the first eigenface, 25% of the second, 4% of the third, and so on.

In optimal conditions (with good lighting, a representative database, front-facing pictures, etc.), a simple eigenface routine might produce accurate readings of up to 90%.

Unfortunately, real life is never that simple, and one must contend with a multitude of 'noisy' factors.

The beauty of this treatment is that even in a large database, each unique face can be expressed very simply using its eigenface decomposition. We no longer have to express each face using thousands of pixels; now, like a simple recipe in which the eigenfaces are the key ingredients, the entire database can be reconstructed as it was before.

A Problem of Distance

Now imagine each face in the database, represented in terms of its eigenface percentages, akin to coordinates lying in some higher-dimensional plane. Our test subject (which may or may not lie in the database) is then projected onto this plane by expressing it in terms of the eigenface components.

Now the problem of recognising the subject becomes as simple as finding the shortest distance (or closest match) between our subject and the faces in the database, a process aided enormously by the fact that each face is now represented by only a handful of eigenface components.

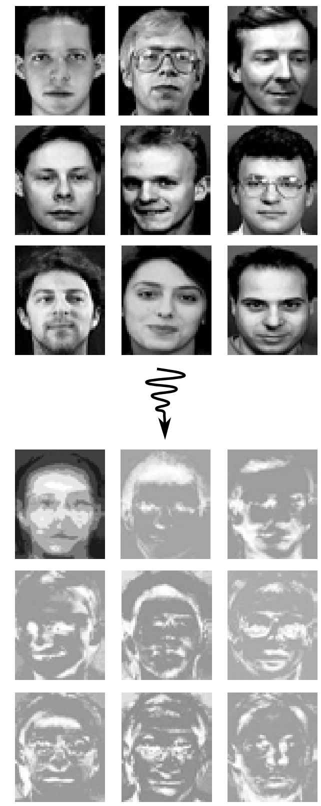


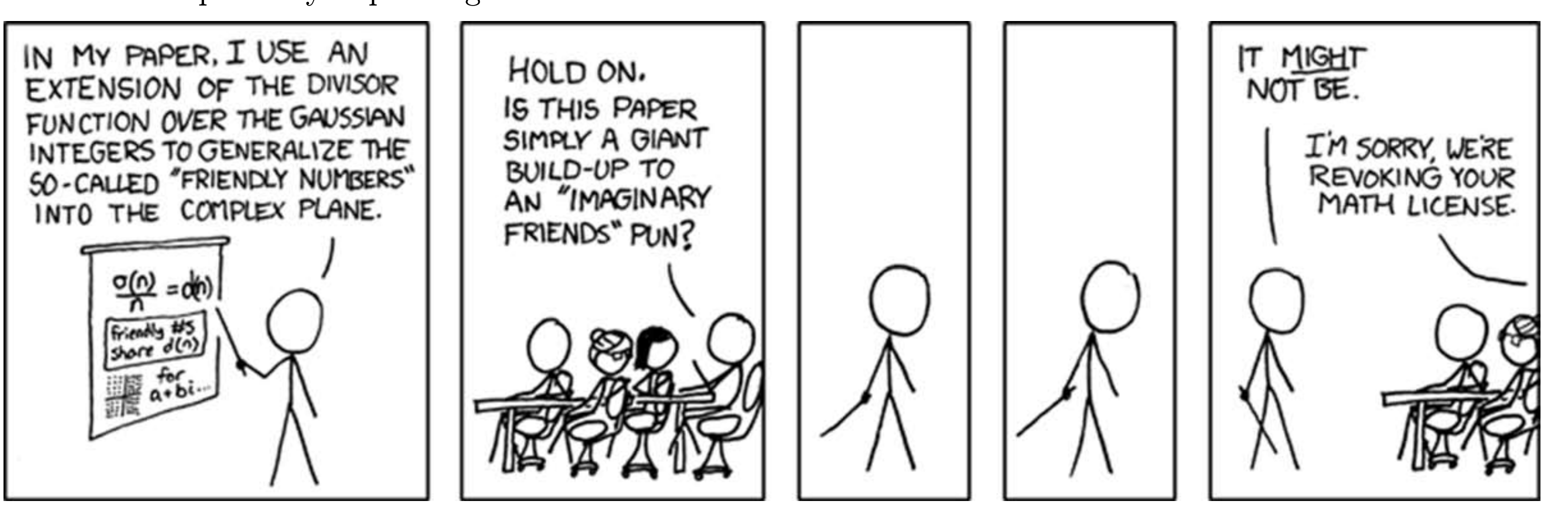
Figure 4: A database of faces can be used to construct an eigenface basis. Afterwards, a face under scrutiny can be decomposed into different percentages of each eigenface (Faces courtesy of AT&T Cambridge)

The Future and You

But really, just how accurate are these eigenface algorithms?

In optimal conditions (with good lighting, a representative database, front-facing pictures, etc.), a simple eigenface routine might produce accurate readings of up to 90%.

Unfortunately, real life is never that simple, and one must contend with a multitude of 'noisy' factors. These include variance in pose (person facing at an angle), obstructions (sunglasses or other people), resolution, lighting, and so on. Despite this, however, the science of facial recognition has steadily improved to the point where today, it is becoming a standard for many military, security, and commercial applications.



http://xkcd.com/

But I digress. You see, the whole point of this article was that the inner workings of facial recognition is nothing but science and mathematics. That's right. It's not half-drunken, hand-wavy speculation that I resemble Brad Pitt and Keanu Reeves. I really, *really* do. It's backed up with science and everything.

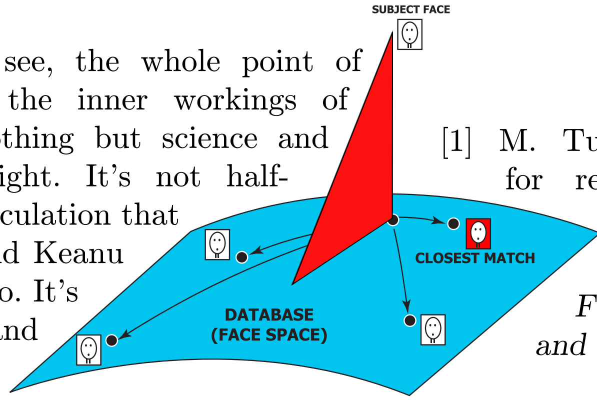
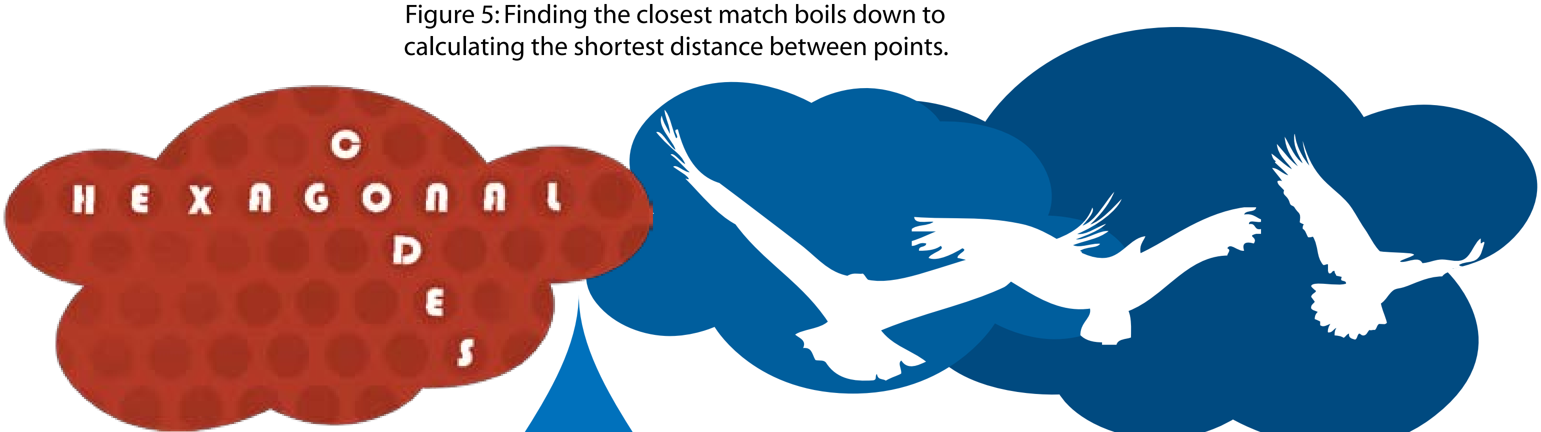


Figure 5: Finding the closest match boils down to calculating the shortest distance between points.

FURTHER READING

[1] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience (1991) 3 (1)*, 71-86.

[2] W, Zhao and R. Chellappa. *Face Processing: Advanced Modeling and Methods*, Academic Press, 2006.



Hexagonal Codes

By
Anthony Quas
and
Peter J. Dukes
University of Victoria

The idea for this article arose from a simple magic trick based on error-correcting codes. This activity has been used in *Math Mania*, an outreach event in which students and staff members at the University of Victoria visit local elementary schools for hands-on math activities with students.

In the trick, the young student is presented with a rectangular array of cups, some face up and some face down. While the magician looks away, the subject is asked to invert (i.e. flip) any cup. Looking back at the cups, the magician then is able to determine which cup the student inverted, without any memorizing! To do this trick, the cups are initially put into a configuration in which each row and each column contains an even number of cups that are the right way up, as in Figure 1. After a cup is inverted, the column in which it sits contains an odd number of cups the right way up (there is either one more or one less than before); and similarly the row in which it sits contains an odd number of cups the right way up. This allows the magician to figure out which row and column the cup that the student inverted is in. Of course this is enough to figure out which cup was inverted.

Figure 1: Rectangular array of cups used in the magic trick.

In this article, we will discuss the connections between this magic trick and error-correcting codes, and investigate an extension where hexagonal arrays replace the rectangular array of cups.

It is shown that the hexagonal pattern corresponds to a more robust code (and hence a flashier magic trick!) More on this later; we now comment on codes and their uses.

In a basic model of a communication system there is a source that produces some sort of data that is to be sent to a receiver. A typical first step in this process is for the data to be transformed into binary digits (***bits*** - 0s and 1s), with strings of several bits representing one piece of data. If we want to use strings of k bits to represent pieces of data then there are 2^k different strings. This means that if there are N different pieces of data, then we need to choose k with $2^k \geq N$ so that all of the possibilities can be encoded. For example, if the data that we want to transmit are letters of the alphabet, then since $2^4 < 26 \leq 2^5$ we need to use

$$\begin{array}{|cccc|}\hline 0&1&1&0\\ 1&0&1&0\\ 0&0&0&0\\ 1&0&0&1\\ 1&0&1&0\\ 1&1&1&1\\ \hline\end{array}\qquad \begin{array}{|cc|c|c|}\hline 0&1&1&0\\ 1&0&1&0\\ 0&0&0&0\\ \hline 1&0&1&1\\ \hline 1&0&1&0\\ 1&1&1&1\\ \hline\end{array}$$

Figure 2: Corresponding array of bits, with one flip

strings of at least 5 bits to encode the letters of the alphabet. We can calculate the number of bits required to be able to encode N different possibilities as $\lceil log_2 N \rceil$(where $\lceil x \rceil$ means x rounded up to the next integer).

If we code letters in binary ('A' = 00001, 'B' = 00010, 'C' = 00011 etc.), then we often run into problems if there is a noisy channel where some of the bits may get corrupted. For instance, consider stray radio interference, a scratch on a CD ROM, etc. Notice that the strings 00110 00001 01100 01100 and 00110 01001 01100 01100 differ in a single bit but represent the words 'FALL' and 'FILL' respectively.

> **Error correction means that you know there was an error and you can work out what was there before the error occurred**

Error-detecting and error-correcting codes are designed to give a way to encode data so that if not too many bits get corrupted, it is possible to detect that there has been an error in transmission or (even better) to correct it. Note that error detection means that you know there has been an error, but you can't tell what the error was whereas error correction means that you know that there was an error and you can work out what was there before the error occurred.

> **Error detection means that there has been an error, but you can't tell what the error was**

Doing this requires that we don't use the minimum possible number of bits to represent the data. The aim is to design error correcting codes which are simple to use, don't involve adding too many extra bits to the strings, and are able to detect and correct most commonly occurring errors.

Codes of this type are used in assigning ISBN numbers to books (so that if someone makes a mistake with one of the digits then computers can detect the error rather than bringing up the wrong book), recording data on CDs (so that if the CD gets some dirt on it then the music can still be read) and in computer memory (so that if an electronic glitch corrupts a memory location, the error can be fixed).

Now let us return to the magic trick with cups. Instead of thinking of cups the right way up and the wrong way up, we write a 1 for a cup the right way and a 0 for a cup the wrong way up. With this view point, the student is presented with an array of bits with an even number of 1s in each row and column. Let's say that a row or column with an even number of 1s (cups the right way up) is called ***balanced***. Inverting the cup corresponds to corrupting a single data bit, thus making some row and column ***unbalanced***. Noting these unbalanced lines is a way of correcting the error. Translating the configuration of Figure 1 into 0s and 1s, we obtain the left array in Figure 2. Flipping the bit in row 4, column 3 leads to the right array in Figure 2. The unbalanced row 4 and column 3 identify the corrupted bit.

Observe that 2 errors cannot always be corrected with this procedure. If the magician sees two odd rows r, r' and two odd columns c, c' he or she is not sure whether positions (r, c) and (r', c') were flipped, or positions (r, c') and (r', c) were flipped. We wonder whether, using a hexagonal array of cups (where there are 3 directions to check in rather than 2), it is possible to accurately correct two (or maybe more) errors. Before discussing the cup trick and hexagonal arrangement further, we give a brief introduction to the theory of error-correcting codes.

> **Codes of this type are used in assigning ISBN numbers to books**

Terminology and facts for codes

A ***word*** of *length* n is a finite sequence of bits $v = b_1 b_2 \ldots\ b_n$, where each bit b_i, is either 0 or 1. A ***code*** is a set of words. The words belonging to some code we are talking about are usually called ***codewords***. We will only be concerned with codes having codewords all of the same length n. This is the ***length*** of the code.

The ***Hamming weight*** of a word v, written wt(v), is the number of times the digit 1 occurs in v. For example, wt(110101) = 4 and wt(00000) = 0.

Let u and v be words of the same length n. The ***Hamming distance*** between them, denoted $d(u, v)$ is the number of positions in which u and v disagree.

If u and v are words of length n, we define $u+v$ to be the word obtained by componentwise addition modulo[1] 2. For example, 01101 + 11001 = 10100.

The minimum distance of a code is important with respect to error detection and correction since it is the minimum number of errors that must occur in order to transform one codeword into another.

You can check that $d(u,v) = \mathrm{wt}(u + v)$, because $u + v$ has a 1 in precisely those positions in which u and v differ. For example, $d(01101, 11001) = 2 = \mathrm{wt}(01101 + 11001) = \mathrm{wt}(10100)$.

It is not hard to see that Hamming distance is symmetric, i.e. $d(u, v) = d(v,u)$, and obeys the so-called ***triangle inequality***: $d(u,w) \leq d(u,v) + d(v,w)$. We use this later without commenting on it - see if you can spot where. For a code C having at least two codewords, the minimum distance of C is the smallest of the numbers $d(v,w)$ over all pairs v,w of distinct codewords in C. For example, let $C = \{0000, 1010, 0111\}$. The minimum distance of C is 2.

The minimum distance of a code is important with respect to error detection and correction since it is the minimum number of errors that must occur in order to transform one codeword into another. A code C can detect all combinations of t or fewer errors if and only if the minimum distance of C is at least $t + 1$. To see this, note that if we start from a codeword and make t errors or less, then we cannot end up at another codeword. This means that we can detect that an error in transmission has taken place. In this case we say that C is ***t-error-detecting***. Even though you can identify that an error has taken place, it may not be possible to fix the error.

We will illustrate error detection with a simple example related to the 'cup game'. Consider starting with all 32 words of length 5. This would be sufficient to encode the English letters, plus some punctuation. For single error detection, we can append a ***check bit*** at the end of each word, creating words of length 6. The rule for the check bit is that the resulting weight must be even. For instance, 00101 becomes 001010 after the check bit (zero) is appended. As another example, 10101 becomes 101011 after the check bit (one) is appended. We will think about the code you get consisting of all words arising this way. In this code, there are still 32 words, each of length 6, and the distance between any pair of different words is at least two. (Verify this for yourself: why can't two codewords, with check bits included, be at distance one?) Consequently, this code is 1-error detecting. To see that it is not error-correcting, imagine that we receive the codeword 001101. There is no way to tell whether it was a corruption of 101101, 011101, 000101, 001001, 001111 or 001100.

The minimum distance of a linear code C is the smallest weight of a non-zero codeword in C.

A stronger condition is that a code C be such that whenever a codeword is subjected to a combination of t or fewer errors, we can guarantee to correctly recover the codeword. For this, we require the minimum distance of C to be at least $2t + 1$. The idea is that if t (or fewer) positions change in v, the resulting 'noise corrupted word' v' is still closer to v than to any other codeword. (The distance from v to v' is at most t so the distance from v' to any other codeword is at least t + 1.) The way to decode a word that is received across a noisy channel is just to replace it by the closest codeword. As we've just seen, provided the code has minimum distance at least $2t + 1$ and no more than t bits change, we are guaranteed to get back to the original word.

The rule for the check bit is that the resulting weight must be even.

Thinking back to the cup game, we think of all the rectangular arrays of 0s and 1s with an even

[1] Addition modulo 2 just means 0 + 1 = 1 + 0 = 1 and 0 + 0 = 1 + 1 = 0

http://twistedpencil.com/

number of 1s in each row and column as the set of codewords in a code (we could write them out as a long string instead of as a rectangular array if we wanted). We pointed out before that the rectangular cup code is 1-error-correcting. It turns out that the minimum distance of the code is 4 (we will see this below) so that the code is 3-error-detecting and 1-error-correcting.

Linear Codes

A code C is called a linear code if $u + v$ is in C whenever both u and v are in C. Observe that u and v are allowed to be the same! In this case, $u + u = 000 \ldots 0$ is in C. Therefore, linear codes always contain the zero word, which we usually abbreviate simply as 0. Structure like this makes linear codes an important branch of coding theory. For example, rather than checking all pairs of words to find the minimum distance, things are a bit simpler for linear codes.

Theorem 1: ***The minimum distance of a linear code C is the smallest weight of a non-zero codeword in C.***

Our strategy to prove this will be the following: if we let m stand for the minimum distance of the code and w stand for the smallest weight of a non-zero codeword, we will show first that $m \leq w$ and secondly that $w \leq m$. Of course this proves that they are equal. This method for showing that two numbers are equal (prove that the first is no bigger than the second; then prove that the second is no bigger than the first) is used all the time in higher mathematics.

Proof. Let u be a word in C of weight w. Since $d(0,u) = w$ and 0 and u are both in the code, we see that the minimum distance of the code is less than or equal to w (i.e. $m \leq w$).

On the other hand, since m is the minimum distance, there must be words v and w in the code with $d(v,\ w) = m$. Since the code is linear, we must have $v + w \in C$. We now have $d(v,\ w) = \mathrm{wt}(v + w)$ from earlier. However, we also have $d(0,\ v + w) = \mathrm{wt}(v + w)$ so that $d(0,\ v + w) = m$. It follows that C contains a codeword $(v + w)$ with weight m so that the minimum weight of a non-zero codeword is less than or equal to m (i.e. $w \leq m$). This shows that $w = m$.

It turns out that the rectangular cup code that we described before is a linear code. We can see this as follows: suppose u and v are two codewords (so that each has an even number of 1s in each row and column). Let's set $w = u + v$ and let's think about any row of w. Suppose that u had $2s$ 1s in the row and v had $2t$ 1s in the row. Let's also assume that there are exactly r places where both u and v have 1s (r can be an even number or an odd number). Now, thinking about how binary addition works, we see that w has a 1 in a spot exactly when u had a 1 there and v had a 0 or vice versa. There are $2s$ - r places in the row where u has a 1 and v has a 0, and $2t$ - r places v had a 1 and u has a 0. This means that the total number of 1s in the chosen row in w is $(2s - r) + (2t - r) = 2(s + t - r)$ - *an even number*. This argument works for every row and every column so that we see that w is in the code. It follows that the code is linear. We now use this together with Theorem 1 to work out the minimum distance of the code. This is the same thing as the minimum weight of a non-zero element.

To be a non-zero element, it must contain a row with at least two 1s in (remember that rows have to contain an even number of 1s). The columns containing the 1s need an even number of 1s in them, so there must be at least one further 1 in each of these columns. This shows that the minimum weight must be at least 4. In fact the minimum weight is exactly 4 as we can think of the configuration with 1s in all four corners and 0s everywhere else. This proves that the minimum distance is 4 as we claimed before.

Every linear code has a ***dimension***, which we now describe. Given a set of words S, the ***linear span*** of S, span(S) is defined to be the set of all words that can be obtained as a sum of elements of S (counting the 0 word as the sum of none of the elements of S).

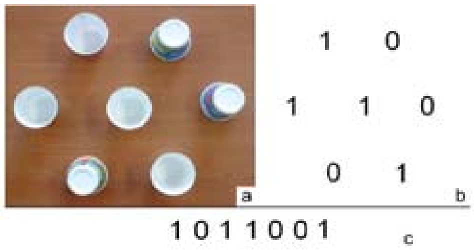


Figure 3: 2-hexagon with associated binary word

Consider the following process applied to a linear code C. First, set $S = \varnothing$. We now keep repeating the following step until span(S) $= C$: Notice that span(S) is a subset of C. If span(S) $\neq C$ we add any element of C that is not in span(S) to S.

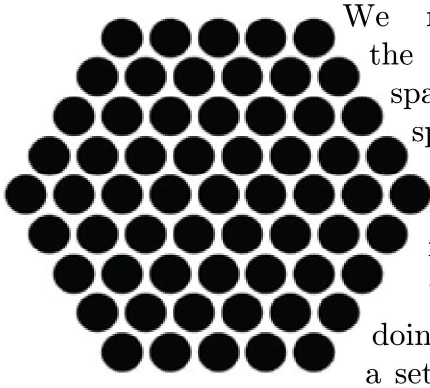


Figure 4: 5-hexagon

When we have finished doing this we end up with a set S with span(S) $= C$. The set S is then called a ***basis*** for C.

For example, suppose our code is $C = \{000, 101, 110, 011\}$. Start with $S = \varnothing$ and add 110 to S. Now span(S) $= \{000, 110\}$. We add 011 to S so that $S = \{110, 011\}$. Now span(S) $= \{000, 110, 011, 101\} = C$ so that $\{110, 011\}$ is a basis for C. Observe that all four words in C are found by summing either none, one of, or both of these basis words. In general, the sum of any subset of the basis words gives a unique codeword in C.

Theorem 2: ***Suppose there are k codewords in a basis for C. Then C has 2^k codewords.***

Corollary: *Every basis for C has the same number of codewords. In other words, the order in which words are crossed off in the process does not affect the number of remaining elements.*

This common number of basis elements for a linear code is what we mean by its ***dimension***. If a linear code with dimension k and length n is used, we say it has ***information rate*** k/n. Thus, in the Example above, C has a dimension 2. This is because if the dimension of the code is k then there are 2^k codewords which means that sending a single codeword carries the same amount of information as k bits. Since the actual length of the word is n bits the information rate is k useful bits per n bits sent.

For example, our earlier code $C = \{000, 101, 110, 011\}$ has information rate 2/3. At one extreme, we have the code {000 … 0,111…1} with one basis element, and information rate $1/n$. At the other extreme, we have the code consisting of all words of length n, having rate $n/n = 1$. But the first code has distance n (and can detect n- 1 errors), while the second code has no error detection at all, since its minimum distance is 1. Both of these codes are rather useless. Coding theory aims for nice compromises between rate and distance.

For further reading on coding theory, consult the web or reference [1].

4 Hex-arrays and correcting two errors

By an ***n-hexagon***, we mean a pattern of cups (or data positions) arranged so that each interior cup is surrounded by six neighbouring cups forming a regular hexagon, and the boundary cups trace out a regular hexagon with n cups per side. For instance, a 1-hexagon is just a single cup, and a 2-hexagon has seven cups as in Figure 3 (a).

The translation from cups to binary words is illustrated in Figures 3 (b) and (c). The array for a 5-hexagon is shown in Figure 4.

By the way, it is a fun exercise to get a formula for the number of cups in an n-hexagon. Notice that there are several geometric differences from the rectangular array. First, the rows go in three directions: horizontal, northwest, and northeast. Also, the number of cups in a line is not constant: it depends on how close the line is to the centre. However, we can still insist that a line is ***balanced*** if the number of cups in it which are the right way up is even.


Figure 5: A weight 6 codeword

Going back to the magic trick, this time on a hexagonal array, the argument we gave earlier shows that the set of codewords (the arrangements that are balanced in each of the three directions)

Figure 6: A 3-hexagon of cups, balanced in every line

is a linear code. In fact, the dimension of this code is the number of cups in an (n-1)-hexagon — see below for more on this. More importantly though, we want to find the minimum distance in the code. By Theorem 1, this is the same as the minimum weight of a non-zero configuration, balanced in every line.

One possible configuration is shown in Figure 5. Check that all lines are balanced, so that the minimum weight in this code is no more than 6. On the other hand, we can see that the minimum weight is at least 6 as follows. In order to be a non-zero configuration, there must be a horizontal line which is not all 0s. If there is just one horizontal line that is not all 0s then we're in trouble because a diagonal line that goes through one of the 1s would have just a single 1 on it (an odd number). This shows that there must be at least two horizontal lines with 1s on them. Similarly for the northwest and northeast directions. If there are 3 or more lines in one of the directions with 1s on then since each line contains at least two 1s, we'd have at least six 1s in total.

The only possibility to rule out is that there are exactly two lines in each of the directions and each line has exactly two 1s so that the total number of 1s in the configuration is 4. Considering the two diagonal directions first, the only possibility would be that the 1s form a parallelogram with 1s at the corners and edges parallel to the two diagonal directions. It is then easy to see that the top and bottom of the parallelogram are in lines with a single 1.

This shows that the minimum weight is at least 6 and so the minimum distance of the code is exactly 6. According to our observations from before, this means that we can detect that up to 5 errors have been made (but not which errors they are), but since $6 \geq 2 \times 2 + 1$, we can correct any 2 errors!

Finally we should do what a good magician should never do and tell you how to perform the magic trick in this case. First you need to set up a hexagonal array of cups so that in each direction there are an even number of cups the right way up. This isn't quite as easy as it sounds but quite a good way to do it is to put the cups in an (n-1)-hexagon in any arrangement at all. It turns out that if you then try to extend the configuration to an n-hexagon, there's always exactly one way to do it. Another way to proceed is to start with all cups down, and repeatedly perform six flips as in Figure 5. (There is also a way to wrap around the boundary by flipping eight cups).

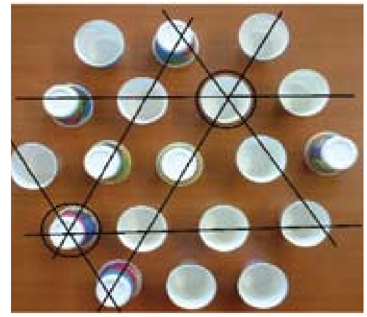


Figure 7: Two flips on different lines.

Next you should invite your victim to turn over any two cups. As the magician you have to figure out which two cups were inverted. Look along the horizontal and diagonal rows and see which rows have odd numbers of upturned cups.

There are 2 basic possibilities. In the first case (see Figure 7) there are two rows in each of the three directions that have odd numbers of upturned cups. In this case there will be exactly two spots where three lines cross. These spots are where the cups were inverted.

In the second case (refer to Figure 8) two cups were inverted on the same line so that this line ended up with an even number of upturned cups. In this case there are two directions that have two unbalanced lines and in the third direction all the lines are balanced. The unbalanced lines taken together trace out a parallelogram. One of the diagonals of the parallelogram is in the third direction. The two corners of the parallelogram on this line are the cups which were inverted.

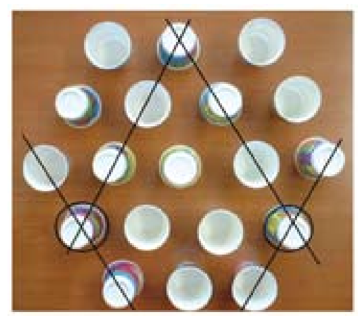
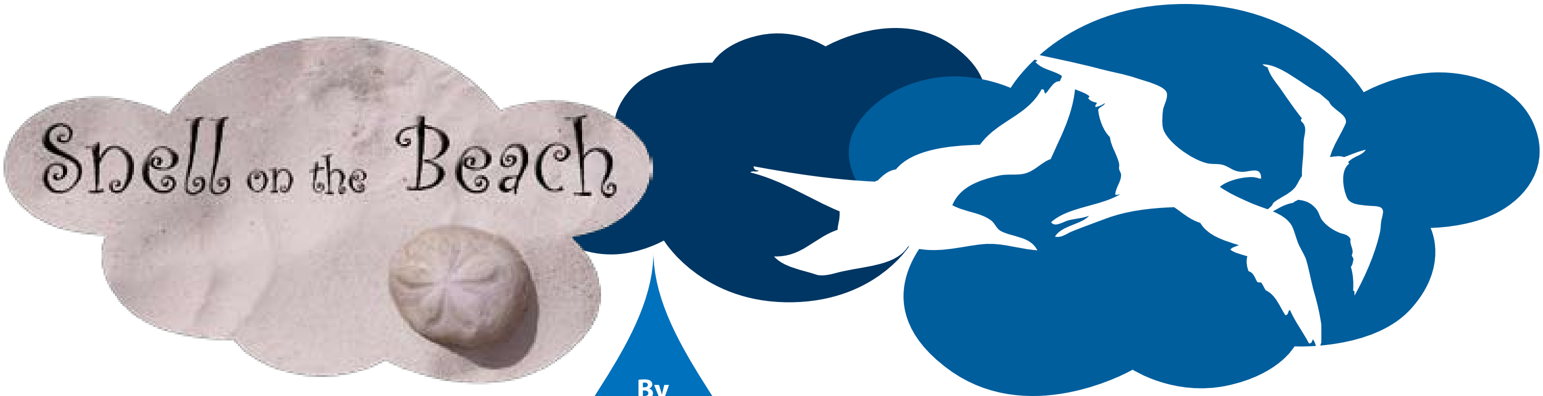


Figure 8: Two flips in the same line.

REFERENCES:

Pless, V. 1982, *Introduction to the Theory of Error-Correcting Code*, John Wiley & Sons Inc.

Snell on the Beach



By
Klaus
Hoechsmann
University of
British Columbia

Carole (C) was sitting on the beach watching her little brother Ivan (I) play in the water. It was only knee-deep, and she was not worried about him at all. He always enjoyed clowning around, grimacing and gesticulating. When he suddenly cried "Help!" Carole just waved at him, but when his head disappeared under water and stayed there, she jumped up and ran to help.

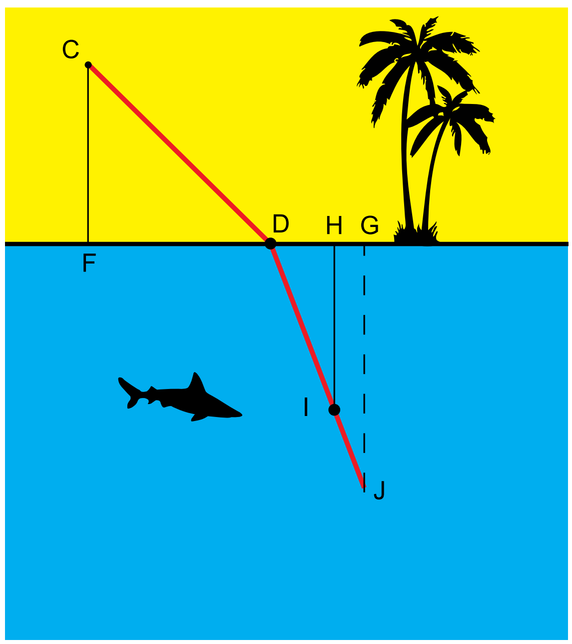


Figure 1

Fred (F) was impressed to see her veer toward Henry (H), almost stepping on Donna (D), instead of going straight for Ivan (see Figure 1). But, Carole was a certified life guard and knew that she would get there faster by this little detour, and Fred knew why this was so.

Since she ran twice as fast on the sand as through the splashing water, she had learned that the cosine of the angle FDC should be twice as big as that of HDI. The general rule was that these cosines should be in the same ratio as the respective speeds. The instructor had said that this was just a fact, called ***Snell's Law***, and that you could not understand it unless you knew Calculus.

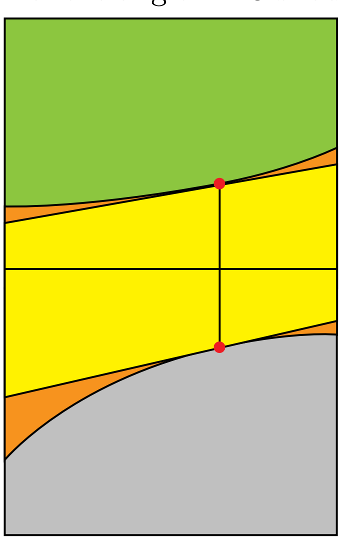


Figure 2

The attentive reader might wonder what Grant (G) and Judy (J) are doing in the picture, and the answer is: *very little*. For now, let us just say that Grant was ogling Judy through his binoculars. This whole story could be told entirely without them, except that Judy stubbornly insisted on being part of this excursion, and as she went so did Grant. We shall say more about them at the end of this tale.

Fred's real name was Fernando, and at school they called him "Ferd the nerd" because he liked mathematics. When Carole had told him about Snell's Law, his first reaction was to pull out a piece of paper and scribble. "It's easy, Carole," he said, ".. but, hm, yes, I *am* using derivatives. There's *got* to be another way: that Dutchman Snell died in 1626 — long before Newton was even born." She was used to that kind of stunt from him, but *this* much detail astonished her. "How come you remember *that*?" she wanted to know. "Ten years after Shakespeare," was his answer.

The next day, he was back with two neat drawings, the first of which is shown here in Figure 2. That was his style: only rarely did he use algebra with Carole because he knew it did not convince her. She had taken an "interdisciplinary" course, where she learned to speak about derivatives and integrals but not to work with them. Fred spent hours with her using reams of graph paper, until one day she exclaimed: "You mean, the steeper the curve, the faster it's moving away from the x-axis? And that *steepness*, is the derivative? Wow!"

Their conversation about the new diagram was rather private, so we'll paraphrase it. If you go

The general rule was that these cosines should be in the same ratio as the respective speeds.

back to the beach picture and imagine D moving from F to H, you'll easily see that CD would be increasing while DI would be decreasing. As you slide your pencil-tip from left to right along the horizontal line in Figure 3, its distance from the upper green area is increasing (like CD), while its distance from the lower gray area is decreasing (like DI). As this drawing is *not to scale* anyway, we might as well regard the lower distance as representing not DI itself but mDI, where m is a positive numerical factor reflecting the ratios of the speeds on sand and through water. Then CD + mDI would be represented by the vertical distance between the green and gray regions, right?

That distance is minimal at the place where they would just barely touch if you slid them together vertically. "Yes," said Carole, "the red points would be kissing." "And the chaperone could separate them with one straight thrust of her cane," Fred added. "They'd share the same tangent line, you mean," Carole sighed, "(you are *so* romantic, Fred) — if you don't wish to slide them together, let us say, it's where their boundaries have the same steepness."

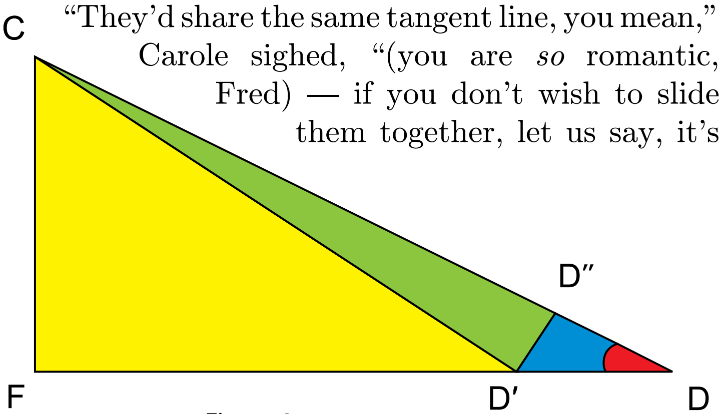


Figure 3

Back to the beach scene. As D′ moves toward D, the triangle FD′C gradually morphs into FDC, and we wish to compare the growth of the hypotenuse (CD′ to CD) with that of the side (FD′ to FD). If we mark D″ on CD so that CD″ always equals CD′ in length (i.e. the green triangle remains isosceles throughout), the changes in hypotenuse and side are represented by D″D and D′D, respectively.

Now, as D′ slides toward D (in synchronisation with its twin D″), the green base angles approach 90 degrees, and the quotient D″D/D′D approaches — hold your breath — the *cosine* of the included angle FDC.

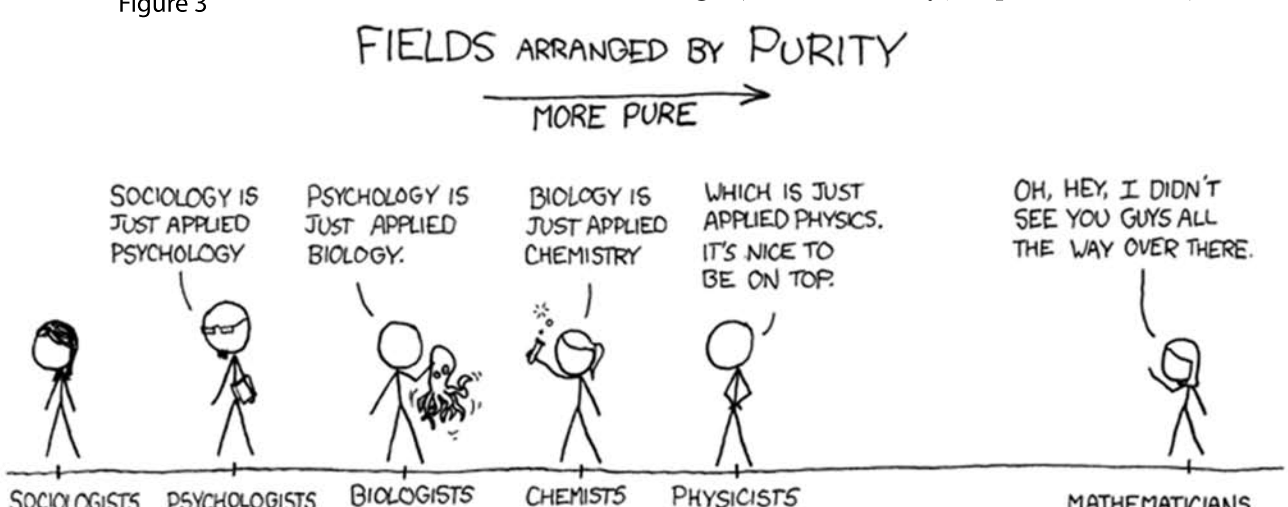
So, now you have all the ingredients to roll your own proof of Snell's Law. But how Grant and Judy fit into this ? Well, it so happens that in issue Number 7 [1] some graphics expert thought it would be pretty if the whole picture were inside a circle, with Donna in the centre and Ivan on the rim (where Judy is now). With CD = DI of equal length, CD + mDI was constant, and the argument was sunk. But who needs a theory when you have facts?

Exercise

If v and w are Carole's speed on land and water, respectively, what is the total time it takes her to reach Ivan? What is the meaning of the factor m? How does it show up in the segment FG?

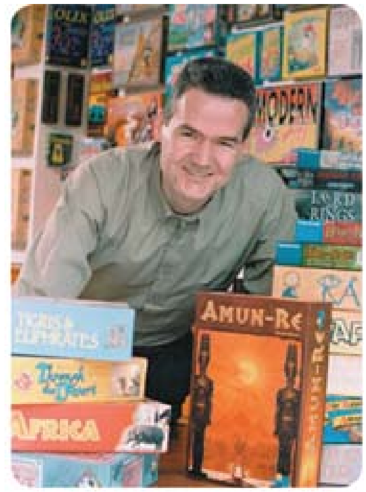
References

[1] Judith V. Grabiner, How Looking for the Best Explanations Revealed the Properties of Light, *Pi in the Sky, September 2003*, 20-22.



http://xkcd.com/

The Board Gaming Genius: Reiner Knizia PhD Mathematics

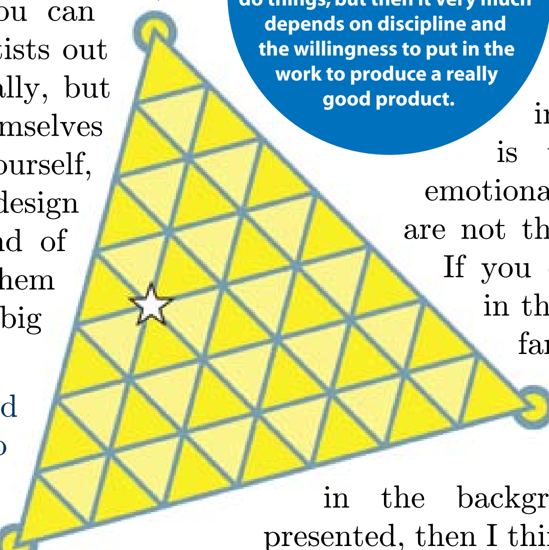


Reiner Knizia has published hundreds of board games including eleven in the top 100 on www.boardgamegeek.com

Interview by Gordon Hamilton

Reiner - I want you to imagine a triangle with the points labeled "Entertainment Businessman," "Creative Artist" and "Scientific Researcher" (researching "What is fun?") Where would you place yourself?

Least on the Scientific Researcher side… and probably in the middle between the Entertainment Businessman and the Creative Artist - because it takes both. I think this is one of the main success criteria in game design: that you can handle both. There are lots of artists out there who are very good artistically, but they are chaotic and don't get themselves organized… and if you organize yourself, but have no artistry (and game design is an art) then you are also kind of empty handed. Getting both of them combined in your life is the big challenge.



"Genius is 1% inspiration and 99% perspiration." Where are you on Thomas Edison's Scale?

I mean 99 and 1 is probably a little bit extreme… I think you actually need both. You need to have a good mind and you need to have the capability to do things, but then it very much depends on discipline and the willingness to put in the work to produce a really good product.

Artists are sometimes called upon to protest with their art. Do you ever feel the urge to protest through your art?

The honest answer is probably 'No.' I think games are there to entertain people, and to bring enjoyment to people. I think there is no mission beyond that. However there are nuances: We have educational games where you engage kids emotionally so that they are open to learning.

Last year, a randomized controlled trial found some games improved mathematical ability [Child Development, March/April 2008, Volume 79, Number 2, Pages 375-394]. Do you think games are under-utilized in mathematics classrooms?

It is difficult to say. I think you can teach math without using games at all, so it's not a necessity; there are many different methods of exciting kids and making them want to learn. I think the important thing in learning is that you have a positive emotional experience; If the emotions are not there the learning is very flat. If you can create positive emotions in the learning process you get by far the best results – and this is where games can help a lot. If the learning content is almost in the background, almost unnoticeably presented, then I think we have an ideal situation. The learning needs to become fun. I think games are an ideal tool for this – they can be used much more widely than today, but they are not the only tool.

Reiner, to avoid becoming biased, you don't play other people's games.

You're right - When you design games you have to take lots and lots of minute decisions. All these decisions have to be taken, and if you already know a decision from somebody else, it is much

“In order to write, you have to live” Ernest Hemingway

more difficult to think of a new one. Once you know how a light bulb works – this is the way to do lights – and if you don't know you might come up with a very different solution…

As a Masters and PhD student did you also protect your mind from bias?

No I did not do that in mathematics – there were many things to learn – and I had great fun in learning them… and I didn't try to re-invent the wheel – so I was quite happy to take in the knowledge from other people and because in mathematics it is not about solving the problem in a different way – once it has been solved it's solved – it's about finding new knowledge – and the more you know, the more new knowledge you can create.

Let's go back to some of your high school experiences with math. Can you think of an instance where you thought “Mathematics is really beautiful!”

I remember looking at mastermind, and working out an algorithm: What do you need to ask, and when you have the answers what do you need to do to actually solve it? I tried to work out the complete algorithm and therefore prove that you could always do it in so many steps. I applied that approach to quite a number of logic puzzles so I was always fascinated by combinatorics – how many different ways are there to do this or that.

The most fascinating questions in mathematics for the general public are probably coming from number theory: “Okay, so I understand what a prime number is. So how do I know that there are an infinite number of prime numbers? How many pairs of prime numbers $[x, x{+}2]$ are there?” You can ask very simple number questions which are unsolved in mathematics and people can really understand. That is rare – most of the questions in mathematics today are complicated and you first need years to understand the question.

I was perplexed that you chose the area of mathematics that you did. You just named combinatorics which was where I would have pegged you. Why not?

I think – apart from finding a mission in life it is extremely important to find good mentors. And one of the great mentors was my professor with whom I did my thesis. He was a professor in Germany and he was a professor in America, and he was working in Analysis so naturally I drifted towards this. I learned a lot from him. With respect to mathematics, but also how to approach problems generally. I could probably have worked in many fields, but I knew him from very early stages in my studies so we kept working together. I did my diploma with him and then my doctoral thesis with him.

Apart from your mentor, who are you standing on the shoulders of – or is game design too young for that?

In game design I probably don't stand on anybody's shoulders. I think I have very much developed my own style and learned by doing and I think game design is a very peculiar profession in a way because you can't really study it.

If you can't study it, and you don't play other people's games - how do you stay fresh?

My teacher always said: “If you don't read your daily newspaper – you will always be uneducated and stupid.” I have always been uneducated and stupid because I've never read it regularly in my life. (*laugh-out-loud*) I have a different philosophy. I'm not interested in day to day news which is relevant today to talk to people and next week is totally irrelevant. I prefer to read books rather than newspapers because they have long term relevance.

It is very important to take in the world. Hemingway said so nicely “In order to write you have to live” and the same is true for game design. Games are a mirror of our times. We have a very hectic life and a small attention span - so the dynamic of the game has to be much much greater today than 50 years ago. Half an hour playing time is almost too long. There are a lot of things that are influenced by our times and I need to take these in if I am to remain successful. So, as well as reading, I stay in contact with people - particularly with the younger age group. I learn what the kids like and then I build that into my design process.

You‘re quoted as saying “Other people steal my ideas before I've ever had them.”

Yes–absolutely!

Is this a big motivator?

If you don't read your daily newspaper - you will always be uneducated and stupid

Yes it is, because I do believe there are lots of games in the universe (like mathematics) and I believe that lots of these games should be mine, and belong to me (I'm kind of smiling when I say this)... and the panic always grabs me when I walk around a fair and I look at games – and I say "hmmm – yes – that's one I should have developed – that's the one I should have found - that should have been mine, I'm too late there and somebody else stole it." I mean this is meant to be funny, but there is a sense of urgency that overcomes me in these fairs...

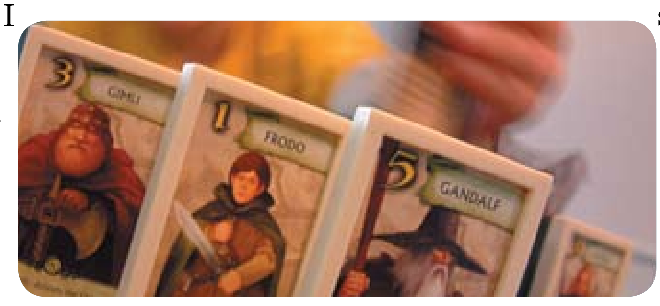
There are the right times for certain things when the technology becomes ripe. For example, I will soon have a game on the market that works with transparent cards... I was very anxious to get it out quickly... I might have been the first one to design a game with transparent cards, but it took a while to get it to market and now I see one or two other products coming out which means I'm only second or third. It's this urgency. When the time is ripe you need to do it and then you need to be fast.

You haven't asked me the most difficult question, which was to explain my PhD thesis to a high school audience.

Well - I subsequently looked at your thesis, and thought that it would be *too* tough a question! ...but go for it if you wish!

Other people steal my ideas before I've ever had them

...There is this funny thing called an integral that we use to calculate the area of certain shapes. Of course we all know very nice formulas that tell us how to calculate the area of specific shapes: The area of a rectangle... length times width; the area of triangle... length times height divided by two. It gets more difficult when you want to calculate the area of a circle – but there is a method there as well – and then you ask, "What happens if I put some holes in there? Can I still have a process to calculate the area?" If I have twenty holes can I do it? If I have a thousand holes can I do it? Yes – because I can still count them, but what if I have uncountably many? Can I still calculate the area? Well then our normal understanding of area starts to fail, but if the process still leads to a result we still have something tangible to talk about. What I worked on in my thesis was a very general mathematical process (which we call the integral) to calculate the "area" for very very crazy shapes where the normal processes that calculate area no longer work.



You did it ;)

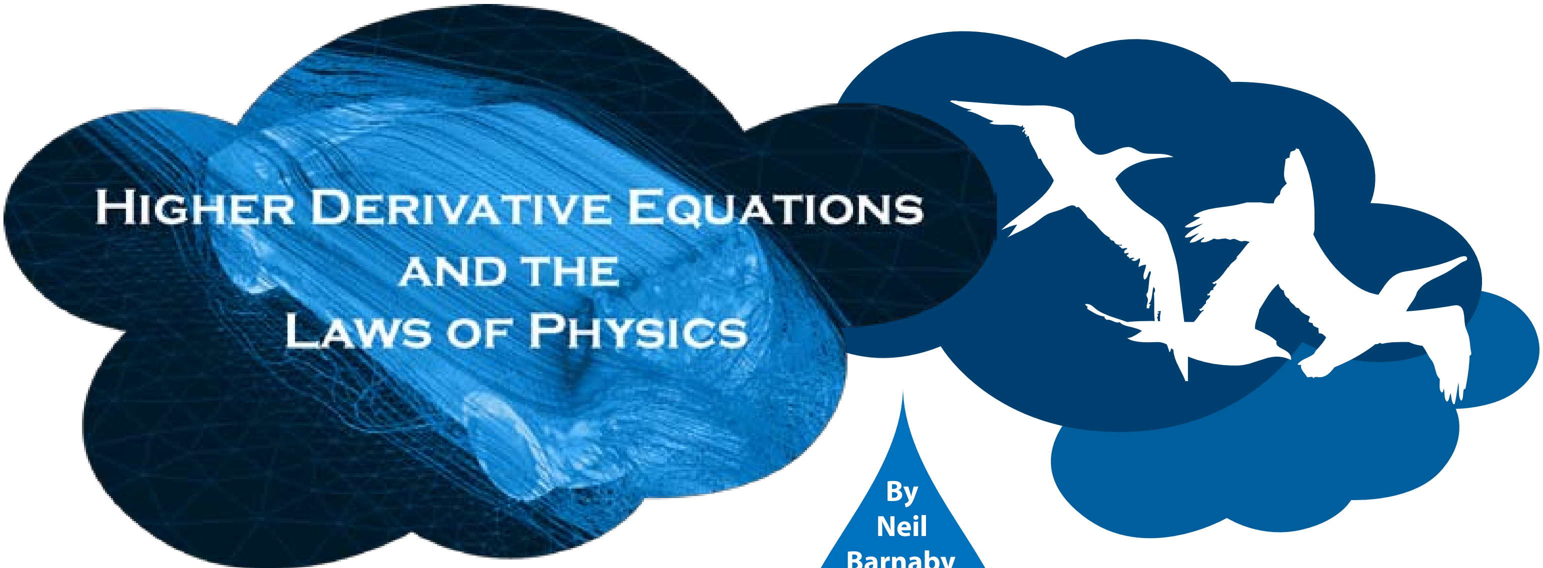
One last quote: "Life offers so many good choices, you can never take them all, and that's good, because it makes life so rich." That's by Reiner Knizia...

Absolutely – that is one of the absolute statements – I don't want to be bored – I think I can only do one or two things in life really properly. There are many other routes I could have taken: I enjoyed being a board member, and running financial companies, and doing mathematical research – I enjoyed all that, and it was a great satisfaction, but I don't think in the end it was my mission.

As a young person, you need to find your mission in life and then to have the courage to follow it. I only found my mission when I got to 40, and I jumped and said "now I do games full time" – I've never regretted it and I've found my niche. I know I miss out on almost everything in life, but as you say "you can have anything, but you can't have everything." Its lots of choices we have to take – and if these are rich choices and you have lots of alternatives then it's a great life.

And this also reflects into games. A good game for me is one that gives me rich choices. I do not want to find the least bad option. I want to have so many good options. I want to sit there biting my finger nails and hoping that when it comes round to me that option is still there.

Higher Derivative Equations and the Laws of Physics



By Neil Barnaby
Canadian Institute for Theoretical Astrophysics

The dynamical laws of physics are described by differential equations. As a theoretical physicist, a big part of the job is solving those equations in various settings and, sometimes, coming up with new equations that might describe nature in some unexplored regime. For me, inventing new theories of physics is the most exciting part of the job. Of course, one can't just start writing down equations haphazardly and expect these to play some role in nature. There are a lot of criteria that we must use to guide our attempts to come up with new physical theories. For example, any candidate for a new law of physics must be studied in detail to make sure that its predictions don't conflict with what is already known about nature. In this article, one of the things I want to accomplish is to give you a sense of what it's like to try to develop new theories of physics.

In this article, one of the things I want to accomplish is to give you a sense of what it's like to try to develop new theories of physics.

As a theoretical physicist, a big part of the job is solving those equations in various settings and, sometimes, coming up with new equations that might describe nature in some unexplored regime.

Before we can try to come up with new equations we need to understand the existing laws of physics. Perhaps the most famous example is Newton's law

$$F = ma\ , \tag{1}$$

that tells us how a body of mass, m, accelerates in response to an applied force, F. Sometimes when we are first introduced to equation (1) it is made to look like a simple algebraic relation, but it really is a differential equation! To see this, suppose that the position x of the mass at any time t is described by the function $x = x(t)$. Velocity is the rate of change of position with respect to time

$$v = \frac{dx}{dt}\ , \tag{2}$$

and acceleration is the rate of change of the velocity

$$a = \frac{dv}{dt} = \frac{d^2x}{dt^2}\ . \tag{3}$$

Using (3) we could replace the acceleration, a, in (1) by a second derivative of position with respect to time. We can also do something similar with the force, F. The particular form of F depends on the system under consideration. In general the force acting on some body might depend on the position, x (for example, restoring force of a spring), or on the velocity, v (for example the viscous drag as a body is moved through molasses). This means that, even without specifying the physical system, we can write Newton's law in true differential form as

$$\frac{d^2x}{dt^2} = \frac{1}{m} F\left[x, \frac{dx}{dt}\right]\ . \tag{4}$$

For any given choice of F we can find a unique solution once the ***initial conditions*** $x(0) = x_0$ and $v(0) = v_0$ are provided. Note that only two initial conditions are needed corresponding to the fact that (4) is second order in derivatives (it contains derivatives no higher that $\frac{d^2}{dt^2}$).

Newton's law is not the only example of a differential equation in physics. There are the Navier-Stokes equations that describe the dynamics of fluids, the Maxwell equations that describe the laws of electro-magnetism, the diffusion equation that describes the propagation of heat through a

http://brownsharpie.courtneygibbons.org/

medium, the Einstein equations that describe how space-time curves in response to matter or energy, the Schrödinger equation that describes the quantum mechanical state of a microscopic system, ... The list could go on and on and on!

I want to draw your attention to a common feature of all these different equations. In *all* known theories of physics, the underlying dynamical equations contain at most two time derivatives when written in terms of fundamental degrees of freedom.[1] It's kind of surprising that so many different equations that describe nature in very different regimes all share such a fundamental mathematical property. In fact, this commonality is *so* fundamental that it is often overlooked. At first glance this seems like a remarkable conspiracy; is there something wrong with higher-order equations that prevents them from playing a fundamental role in physics?

In all known theories of physics, the underlying dynamical equations contain at most two time derivatives when written in terms of fundamental degrees of freedom.

To answer this question, let's consider an explicit example. Suppose we focus on a mass m bouncing around on a spring. We denote the displacement of the mass away from the equilibrium position by $x(t)$. As long as $x(t)$ is not too large, then the restoring force is proportional to the displacement

$$F \approx -k\,x(t)\ , \tag{5}$$

where k is a constant that characterizes the stiffness of the spring. (Don't let the minus sign in this equation bother you, it simply reminds us that the force is always pulling the mass back towards x=0.) We have written equation (5) as an approximation because at larger displacements there will be nonlinear corrections like αx^3, βx^5, etc. Newton's law for this system now takes the simple form

$$\frac{d^2x}{dt^2} + \omega^2 x = 0\ , \tag{6}$$

where we have defined $\omega = \sqrt{k/m}$ in order to make the equation look simpler. The general solution of equation (6) is

$$x(t) = x_0\cos(\omega t) + \frac{v_0}{\omega}\sin(\omega t)\ . \tag{7}$$

The solution (7) describes a function that oscillates in time with frequency ω, just the behaviour we expect a mass on the end of a spring! This function is illustrated in Figure 1 for the special case where mass is released from rest so that v_0=0.

Exercise

Check that (7) really does solve (6) and also that x_0, v_0 appear correctly to be interpreted as initial position and velocity.

Often in physics one would like to know the total energy of a system. There are two kinds of energy: kinetic and potential. Kinetic energy is generically given by

$$K = \frac{1}{2}mv^2\ . \tag{8}$$

For our simple harmonic oscillator the potential energy is

$$U = \frac{1}{2}m\omega^2x^2\ . \tag{9}$$

The total energy is given by the sum of the potential and kinetic energies

$$E = K + U = \frac{1}{2}mv^2 + \frac{1}{2}m\omega^2x^2\ . \tag{10}$$

To check that this really is the right expression, notice that E defined by (10) is conserved. That is, if you take the derivative you'll find that

$$\frac{dE}{dt} = 0 \tag{11}$$

[1] There is a caveat here that you could obtain higher order equations by taking derivatives or by decoupling a set of equations. I'm referring only to the most fundamental formulation. So I'm definitely *not* trying to claim that higher order equations are never useful in physics, only that where they appear it is as a stand-in for a more fundamental set of lower-order equations.

implying that E is a constant in time.

Exercise

Check that E really is conserved by taking the derivative of (10) and using equation (6).

Now, let's see what would happen if we decided to modify Newton's laws by adding some higher derivative terms. This might sound like a funny thing to do (after all, we know that Newton's law works perfectly fine!) but it's a worthwhile exercise. Often, we can learn a lot about *why* the laws of nature take the form they do by imagining how things would be different if they took some other form. So, let's boldly consider replacing (4) by the following equation

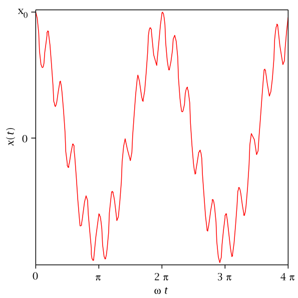


Figure 2: A plot of the solution (14).

$$g\frac{d^4x}{dt^4} + \frac{d^2x}{dt^2} = \frac{1}{m}F\left[x, \frac{dx}{dt}\right] , \tag{12}$$

where g is a new constant of nature having dimensions of seconds squared. If we take g=0 then obviously this reproduces the usual force law (4). For g sufficiently small we expect that equation (12) will give basically the same predictions as (4). So (12) seems like a decent starting point for our first attempt at modifying the laws of nature.

To make life even simpler, let's focus again on our mass-on-a-spring example. Using the restoring force (5) with our modified Newton's law (12) we find

$$g\frac{d^4x}{dt^4} + \frac{d^2x}{dt^2} + \omega^2 x = 0 . \tag{13}$$

This equation is not much harder to solve than (6). The general solution looks like

$$\begin{aligned} x(t) &= A_+\cos(\omega_+ t) + B_+\sin(\omega_+ t) \\ &+ A_-\cos(\omega_- t) + B_-\sin(\omega_- t) , \end{aligned} \tag{14}$$

where we have defined the frequencies

$$\omega_\pm = \frac{1}{\sqrt{2g}}\sqrt{1 \pm \sqrt{1-4\omega^2 g}} . \tag{15}$$

The solution contains four arbitrary constants A±, B± rather than two because (12) is fourth order in time derivatives. These constants could be related to the initial conditions $x(0)$, $v(0)$, etc. But we don't need to go through that effort to make our point. The solution (14) is plotted in Figure 2 with $g\omega^2 = 0.01$ and a representative choice of initial conditions. Compare this plot to Figure 1.

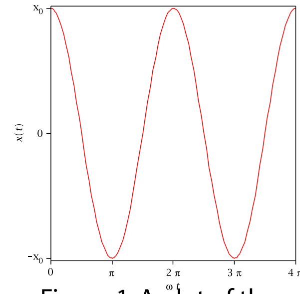


Figure 1: A plot of the solution (7).

Exercise

Verify that (14) solves equation (13).

To see what's wrong with equation (13), we need to compute the energy. It turns out in our higher derivative theory equation (9) still gives the potential energy, however, the simple expression (8) for the kinetic energy is no longer valid. The easiest way to see that is by noting that the total energy (10) will no longer be conserved if the motion of $x(t)$ is described by equation (13) rather than (6). So we need a new expression for the kinetic energy. The derivation is pretty complicated and involves a lot of theoretical machinery beyond the high school level. Let me simply tell you the answer:

$$\begin{aligned} K &= mg\frac{dx}{dt}\frac{d^3x}{dt^3} - \frac{mg}{2}\left(\frac{d^2x}{dt^2}\right)^2 \\ &+ \frac{m}{2}\left(\frac{dx}{dt}\right)^2 . \end{aligned} \tag{16}$$

Using this new expression for kinetic energy the total energy E=K+U will be conserved if you use equation (13) to describe the dynamics of $x(t)$, just as it should be.

Often, we can learn a lot about why the laws of nature take the form they do by imagining how things would be different if they took some other form.

Aside from being very complicated, there's something deeply troubling about equation (16): the kinetic energy can be negative! Regardless of the sign of g we can always imagine some function $x(t)$ that makes this expression less than zero. This is unlike any system that we are used to in physics. At first glace it's not at all obvious how to make sense of negative kinetic energies. In physics, negative kinetic energies are usually interpreted as signaling an instability.[2] Physically this is unacceptable: we

[2] We didn't really see this instability manifest itself in our toy model because we didn't include nonlinear terms like αx^3 in equation (5).

know that a mass bobbing around on the end of a spring is a completely stable system.

> In physics, negative kinetic energies are usually interpreted as signaling an instability. Physically this is unacceptable: we know that a mass bobbing around on the end of a spring is a completely stable system.

It's important to be clear about exactly what the problem with equation (13) is. This is a perfectly well-defined mathematical equation. However, it isn't an acceptable law of physics because it predicts something - namely negative kinetic energy - that is not compatible with the world we see around us. So we are forced to discard our first attempt to modify Newton's law and regard equation (13) as a dead end. But what about more complicated equations? What if we added even more derivatives? In 1850 a Russian mathematician named Ostrogradski was able to prove an amazingly general theorem [1]. Ostrogradski showed that *any* fundamental laws of physics involving n^{th} order differential equations will lead to negative kinetic energy if $n > 2$. This powerful theorem precludes the possibility of higher derivative equations playing any fundamental role in physics. Moreover, it explains why every successful law of physics that has been discovered to date has involved no more than two time derivatives.

> Ostrogradski showed that any fundamental laws of physics involving N-th order differential equations will lead to negative kinetic energy if N > 2.

> This powerful theorem precludes the possibility of higher derivative equations playing any fundamental role in physics. Moreover, it explains why every successful law of physics that has been discovered to date has involved no more than two time derivatives

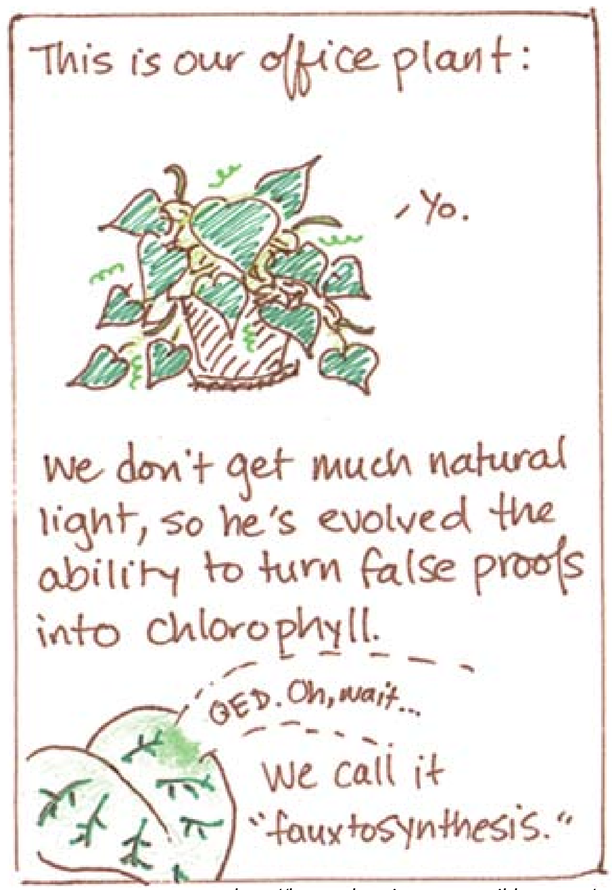
Ostrogradksi's theorem is very general and also very deep. It explains what might otherwise have looked like a strange coincidence and it also serves as a guide for trying to modify the laws of nature. However, this theorem is not widely appreciated by most practicing physicists. The reason, I think, is that this theorem is so fundamental. Most of us take it for granted that the laws of nature must be second order in derivatives without ever stopping to ask why.

You might find it a little depressing that our first attempt to modify the laws of physics was unsuccessful. But don't despair! Indeed, many new theories of physics are found to be incompatible with nature even without the need to perform any new experiments. Failed theories are still useful because they help to guide future attempts. It turns out that there is a loophole to Ostrogradski's theorem that allows us to add higher derivatives to the laws of nature without running into trouble with instabilities. However, it requires recourse to some exotic mathematical objects (for example equations with *infinitely* many derivatives) and a complete discussion is beyond the scope of this article.

> Failed theories are still useful because they help to guide future attempts.

References

[1] M. Ostrogradksi, *Mem. Ac. St. Petersbourg* **VI 4** 385 (1850).



http://brownsharpie.courtneygibbons.org/

Domenico di Michelino, *La Divina Commedia di Dante* (*Dante and the Divine Comedy*). Fresco in the nave of the Duomo of Florence, Italy - the terrace purgatorio lies over his right shoulder

By Gordon Hamilton
Masters Academy & College

Since his death in 1968, George Gamow has been living in purgatory in a seven floor terraced apartment building with penthouse garden on top. Scummy tenants reside on the lowest floor, but the residents become increasingly sophisticated as one moves up. George's apartment is on the third floor.

Each day George wakes up, dresses, eats, locks-up, walks down the corridor to the elevator and impetuously pushes both the up and down buttons. When the elevator arrives – he compulsively steps in – before checking whether it's ascending from the nicotine miasma of the lower floors or descending from the perfumed fragrances of the upper floors.

If it's ascending from the lower floors, George gets a nasty elevator ride, and is off to a bad start in the day.

What goes up must come down - so the first impulse is to think that George is going to have a bad start to the day about ½ the time. Is that right or wrong? Before we try to solve the problem, we need to understand exactly how the elevator operates:

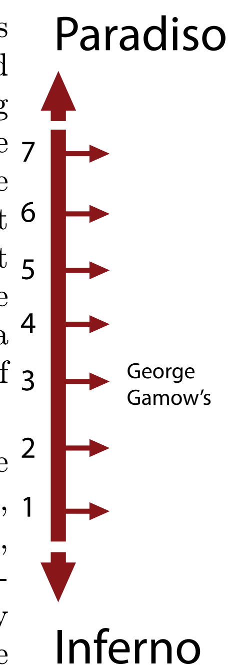
"...the elevator almost always appeared as it was going down, and so the journey [up to see his friend] would be a frustrating 'down and then up'..."

Julian Havil in "Impossible?" (reviewed in Pi in the Sky - #12)

- The elevator goes up and down picking up and dropping off souls on different floors. The elevator's destination is always randomly determined - so when the elevator is at rest, it is equally likely to be on any floor.
- It goes at a constant speed when moving, and when it has no requests, it just waits where it is.
- When the elevator is going past a floor, souls like George can jump on or off without altering the final destination of the elevator.

In Julian Havil's "Impossible?" (reviewed in Volume 12) George's original calculations are extended to include multiple elevators. Although he does not state it, Julian's multiple elevators are assumed stationary when George pushes the buttons.

In his 1958 book, *Puzzle Math*, George showed the impulsive answer of ½ is wrong. He argued that the elevator's last stop before visiting his floor was equally likely to be 1 or 2 (below), or 4, 5, 6 or 7 (above) so the elevator is twice as likely to have descended from a sweet smelling floor than to have ascended from the stench below. That was great news. It meant that George could expect to have a bad elevator ride only 1/3rd of the time.



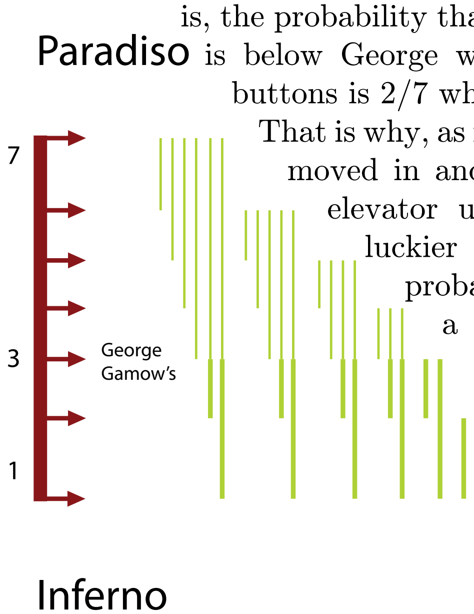
Sure enough, when George arrived in purgatory in 1968, before the population explosion, he confirmed that his day-to-day elevator rides were a nasty experience only about 1/3rd of the time.

But over the last few years, George has realized that something is wonderfully wrong. He is now getting bad elevator experiences a lot less frequently than 1 in 3 and his luck is especially good during morning rush-hour. Find an explanation for George's statistical joy. (Spoiler alert: answer on the next page.)

George Gamow
1904 – 1968

The only thing that is different now compared to 1968 is that there are more souls after the population explosion. The only thing different between rush-hour and non-rush-hour is that there are more souls using the elevator at rush-hour. Using these two clues, George finally figures out that before, when he pushed the buttons, the elevator was at rest. The probability was 1/3rd of a bad elevator experience. Now, during rush-hour, the elevators are moving when George pushes the buttons. Does that make a difference? The elevator travels with a constant speed, so all we need to do is to find all possible origin-destination pairs and find out how much of each is below George's floor.

The total length beneath George Gamow's is 16 floor heights out of a total of $21 + 15 + 10 + 6 + 3 + 1 = 56$ floor heights. That is, the probability that a moving elevator is below George when he presses the buttons is $2/7$ which is less than $1/3$. That is why, as more and more souls moved in and started increasing elevator usage – George got luckier and luckier as his probability of getting a smelly elevator dropped from $1/3$ to $2/7$.

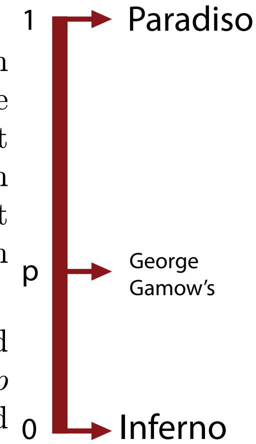


In general, if George is on floor f in a building with F floors: The total length of possible floor-to-floor commutes is a tetrahedral number: $(F - 1)(F)(F + 1)/6$ and the total length of these below George's floor is another tetrahedral number $(f - 1)(f)(f + 1)/6$ plus a "triangularprism" number $(F - f)(f - 1)(f)/2$.

Probability (elevator below)

$$= \frac{\dfrac{(f-1)(f)(f+1)}{6} + \dfrac{(F-f)(f-1)(f)}{2}}{\dfrac{(F-1)(F)(F+1)}{6}}$$

In an attempt to deal with the purgatory population explosion (which lags behind the earth's population explosion by an average lifespan), the transcendental administration decided last month to scrap the outdated terraced landscaping that Dante visited, and replace it with a modern apartment complex that has each soul allocated to their own unique floor.



Earlier this week, George learned that he will be allocated floor p where $0 \leq p \leq 1$. He quickly figured out that his probability of having a smelly elevator in the morning is:

- p if he presses the buttons when the elevator is stationary.
- $p^2(3\text{-}2p)$ if he presses the buttons when the elevator is moving. George used integration to find this, but he might also have used the limit of the ratio above:

Probability (elevator below)

$$= \lim_{F\to\infty} \frac{\dfrac{(pF-1)(pF)(pF+1)}{6} + \dfrac{(F-pF)(pF-1)(pF)}{2}}{\dfrac{(F-1)(F)(F+1)}{6}}$$

$$= \lim_{F\to\infty} \frac{(pF-1)(pF)(pF+1) + 3(F-pF)(pF-1)(pF)}{(F-1)(F)(F+1)}$$

$$= \frac{(pF)^3 + 3(F-pF)(pF)^2}{(F)^3}$$

$$= p^2(3-2p)$$

http://twistedpencil.com/

Sacred Mathematics

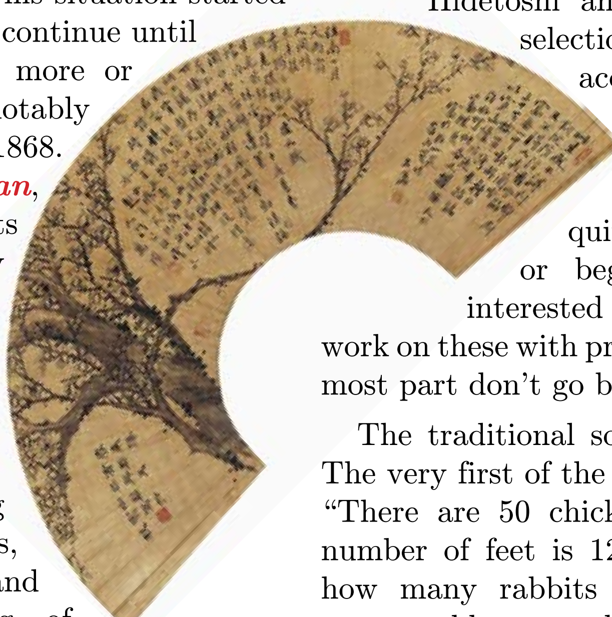
Japanese Temple Geometry

By Fukagawa Hidetoshi and Tony Rothman
Princeton University Press
ISBN 978-0-691-12745-3

Reviewed by Tom Archibald
Simon Fraser University

We usually think of mathematics as something that is largely independent of an individual culture. However, when a society is isolated from the rest of the world, mathematical thought will continue and sometimes develops in a direction that is independent of what is happening elsewhere. That is exactly what took place in Japan in the Edo period, when the feudal government of the Tokugawa clan decreed that Japan should be closed to the outside world. This situation started in the early 1600s and was to continue until the country was opened up more or less forcibly to the west, notably with the Meiji restoration of 1868. During that long period, ***wasan***, Japanese mathematics, went its own way. This has been very nicely discussed in an excellent book by Annick Horiuchi, but you have to read French. [1]

The present book, in English, also discusses this period, but rather than looking at research mathematics, it presents a remarkable and intriguing fad: the posting of mathematical problems, particularly in geometry, on pieces of wood outside temples as a kind of offering. This is the "sacred mathematics" of the book's title. Offerings of images at Japanese temples became very common in the Edo period, and were usually painted on wooden shingles in ink. The most common images were lucky animals, such as horses, but the images associated with mathematical problems, called ***sangaku***, filled the dual role of showing reverence to the god and demonstrating the mathematical prowess of the individual who made the offering. The problems were also collected in books and apparently were widely popular, and there are also manuscripts containing such problems that have not been found posted at temples. The problems range from easy to very difficult.



Hidetoshi and Rothman here present a selection of the problems, graded according to difficulty, with hints and solutions. The aim is to give the flavour of these works, and this is quite successful. High school or beginning university students interested in problems can read and work on these with profit, since the methods for the most part don't go beyond elementary calculus.

The traditional solutions are often surprising. The very first of the ninety problems cited states: "There are 50 chickens and rabbits. The total number of feet is 122. How many chickens and how many rabbits are there?" Most students now would approach this with algebra, but the traditional solution begins: "If rabbits were chickens the total number of feet would be 100 ..." A problem that has a more Japanese flavour involves drawing figures on a folding fan. Unfolded, the Japanese fan is a sector of a circular annulus or ring. Suppose the sector has radius R, and draw

[1] A. Horiuchi, *les mathématiques japonaises à l'époque d'Edo*. Paris: Vrin, 1994.

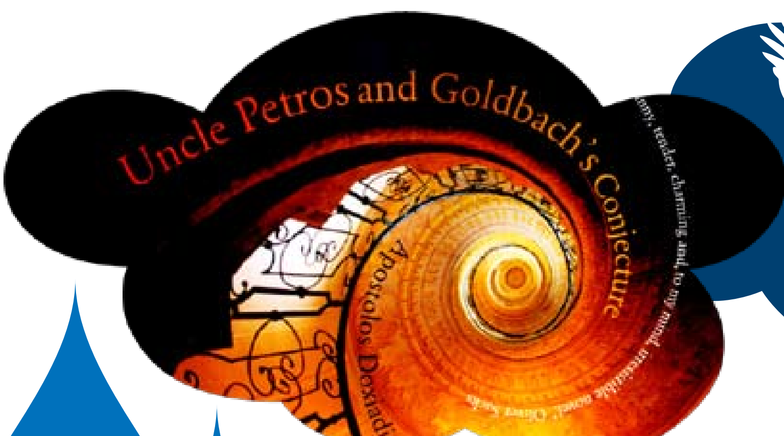
two circles of the same radius r that touch one another and also the outer and inner edges of the fan. we draw a chord length d to the outer edge tangent to the tops of the two circles, and add a small circle radius t tangent to the chord and the fan, as shown. Given d and the diameter of the small circle, we are invited to find the diameter of the large circles. The proposer gives the answer (3.025 if d=3.62438 and $42t$=3.025) but not the method. Indeed, we don't know his method, since the answer is incorrect. The authors don't provide an explanation for the error; nor do they suggest any reasons for these rather surprising values. But they do give a solution, involving constructing some auxiliary lines and using the Pythagorean theorem.

There are 50 chickens and rabbits. The total number of feet is 122. How many chickens and how many rabbits are there?

The authors also give some basic historical background and examples of the Chinese sources of Japanese mathematical work. Hence the book gives a nice glimpse of East Asian mathematics of several centuries past. Not all of the historical details are as carefully written as would be ideal, for example on the subject of the language of the tablets. This doesn't detract from the interest and overall usefulness of the volume, especially for the student reader, and it should serve to stimulate further reading on the mathematics of these cultures. The illustrations, some in full colour, will also entice the reader.

Most students now would approach this with algebra, but the traditional solution begins: "If rabbits were chickens the total number of feet would be 100 ...

Students and teachers will read and work on the problems in this book with pleasure. They give an interesting glimpse of another culture, one in which the value of mathematical problem solving merits a kind of religious recognition. An inscription on one of the ***sangaku*** gives a hint of why: "Confucius says, you should devote all of your time to study, forgetting to have meals and going without sleep. His words are precious to us. Since I was a boy I have been studying mathematics and have read many books on mathematics. When I had questions, I visited and asked mathematician Ono Eijya. I appreciate my Master's teachings. For his kindness, I will hang a ***sangaku*** in this temple." These days, such recognitions are unusual.



By Apostolas Doxiadis
Faber and Faber
ISBN-13 978-0571205110

Reviewed by Clara Park
CheongShim International Academy

Goldbach's conjecture is easy to state:

"***Every even number greater than 2 is the sum of two primes.***"

This statement seems to be straightforward as we can attempt proving it straight away.

4=2+2

6=3+3

8=3+5

10=3+7=5+5

12=5+7

14=3+11=7+7...

We can conclude easily that this theorem is plausible! But, seemingly, this simple theorem, known as Goldbach's Conjecture, has remained one of the most tormenting problems in the field of mathematics, along with Fermat's Last Theorem. For centuries, many talented mathematicians were enticed to this myth but were unable to prove the theorem.

It will appeal to almost any student, even students with no previous interest or background in mathematics.

Thanks to today's technology, mathematicians discovered, with the help of supercomputers, that the conjecture holds true for all even numbers up to 400,000,000,000,000 (four hundred trillion). Faced with the reality that "numbers are infinite," however, they were once again depressed since they are still not able to provide a general proof.

Books on mathematics are often 'how-to' guides on problem-solving or earnest treatises on some area of mathematics, but Apostolas Doxiadis' *Uncle Petros and Goldbach's Conjecture* takes the form of a fictional story that can entertain and inform a wide range of readers regardless of their mathematical knowledge. Readers may feel the same sort of enjoyment as watching a movie, such as *Forrest Gump*.

The main character, Petros Papachristos, is a mathematical genius but considered a complete failure in life. Recklessly, he abandoned a successful career in academics, shutting himself up away from family and friends in order to devote his every waking hour to proving Goldbach's conjecture, a theorem that mathematicians have been unable to solve for almost three centuries. As foolish as any of us might believe such an action might be, through the eyes of an understanding nephew, we, too, soon come to understand Petros' passion and realize he is deserving of our respect. We empathize with Petros' bleak hope of solving the problem, and we are just as exhilarated as Petros' nephew when he learns that his uncle has finally cracked the secret.

The story is not about problem solving. It is rather a story of challenge, persistence, and devotion. Petros' obsession with the problem suggests that the result is not the only thing that matters in life.

The story is not about problem solving. It is rather a story of challenge, persistence, and devotion.

I strongly recommend this book to other readers. It will appeal to almost any student, even students with no previous interest or background in mathematics. They will be stimulated and entertained as well as delighting simply in the spirit of determination of all those people, including mathematicians, who risk all in attempting to do the seemingly impossible.

Math Challange Winner Announced

Obtaining solutions to four Math Challenge Problems posed in Issue #11 (Spring 2008) proved challenging to our readers. At the time of publication of Issue #12 (Fall 2008) we had not received any solutions, so the deadline for submissions was extended to February 28, 2009. I am delighted to announce the decision of our panel, who decided to split the prize between two very talented students. Each submitted three correct solutions to the Challenge Problems, however, neither submitted a solution to the fourth one - Problem #6.

The two winners, who will split the $300 prize money, are: Joshua Lam of The Leys School, Cambridge, England and Rati Gelashvili, of Tbilisi, Georgia, currently studying at the Georgian Technical University. Congratulations to Joshua and Rati.

Solutions to the four Challenge Problems appear in this Issue of Pi in the Sky. We have included some solutions by our two Prize winners.

SOLUTIONS

Math Challenges Problems 4, 5, 6 and 7 Published in Issue 11 (Spring 2008) of Pi in the Sky

4 Prove that any positive integer n coprime to 10 is a divisor of a repunit.

Solution by Rati Gelashvili

Since $(9n, 10) = 1$ according to Euler-Fermat's theorem $9n$ is a divisor of $10^{\varphi(9n)}$ - 1 = 99...99. Consequently, n is a divisor of 11...11.

Alternative Solution by Joshua Lam

Let m be a positive integer coprime to 10: Let R_n be the repunit with n 1s. Consider each R_n (mod m). Since there are infinitely many R_n, but only m different residues (mod m); by the Pigeonhole Principle we can find $i > j$ such that $R_i \equiv R_j$ (mod m). Then R_i - $R_j \equiv 0$ (mod m), hence

m | $R_{i-j} \cdot 10^j$ and since $(m, 10) = 1$ we conclude that m | R_{i-j}

Also solved by Simon Morris

6 R and r are respectively the radii of the spheres circumscribed about, and inscribed in a tetrahedron. Prove that $R \geq 3r$.

Solution

Let us consider the tetrahedron having the vertices at the centroids of the faces of the given tetrahedron. This tetrahedron is similar to the given one and the ratio of similitude is 1/3 . Its circumscribed sphere is therefore of radius $R/3$. This sphere intersects all the faces of the tetrahedron, therefore its radius should be greater or equal than the radius of the inscribed sphere.

5 Let n be a positive integer and let φ(n) denote the number of positive integers less or equal to n that are coprime to n. Prove that for any positive integers m and n

$$\varphi(mn) \leq \sqrt{\varphi(m^2)\varphi(n^2)}.$$

Solution

Let p_i be the prime factors of m and n, q_j the prime factors of m only, and r_k the prime factors of n only.

We have

$$\varphi(mn) = mn\prod_i\left(1-\frac{1}{p_i}\right)\prod_j\left(1-\frac{1}{q_j}\right)\prod_k\left(1-\frac{1}{r_k}\right)$$

$$\varphi(m^2) = m^2\prod_i\left(1-\frac{1}{p_i}\right)\prod_j\left(1-\frac{1}{q_j}\right)$$

$$\varphi(n^2) = n^2\prod_i\left(1-\frac{1}{p_i}\right)\prod_k\left(1-\frac{1}{r_k}\right)$$

Notice that each product $\prod_i$, $\prod_j$ or $\prod_k$ is replaced by 1 if there are no factors containing p_i, q_j, nor r_k.

Now the required inequality is equivalent to

$$\prod_j\left(1-\frac{1}{q_j}\right)\prod_k\left(1-\frac{1}{r_k}\right) \leq 1$$

that is obviously true. The equality occurs only if m and n have the same prime factors.

A similar solution was submitted by Joshua Lam and Rati Gelashvili

7

Prove that in any convex polyhedron (a) there is either a triangle face or a vertex at which three edges meet and (b) there is a face having less then six sides.

Solution

If v, e, f denotes the number of vertices, edges and respectively faces of the polyhedron then we have Euler's polyhedral formula

$v - e + f = 2$.

Let $F_1, \ldots, F_f, V_1, \ldots V_v$ be the faces and the vertices of the polyhedron, f_i the number of edges of F_i and v_i the number of edges issuing from V_i.

(a) Assume the contrary, that there is no triangle face and there is no vertex at which three edges meet, i.e.

$f_i \geq 4$, $v_i \geq 4$

Then we have:

$$2e = \sum_{i=1}^{f} f_i \geq \sum_{i=1}^{f} 4 = 4f \text{ and } 2e = \sum_{i=1}^{v} v_i \geq \sum_{i=1}^{v} 4 = 4v$$

hence $e \geq 2f$ and $e \geq 2v$.

By using these inequalities and Euler's polyhedral formula we get

$4 = (2f - e) + (2v - e) \leq 0$

which is a contradiction. Therefore any convex polyhedron should have at least a face with three edges or at least a vertex at which three edges meet.

(b) Assume by contradiction that every face has more then five edges. As above we have:

$$2e = \sum_{i=1}^{f} f_i \geq \sum_{i=1}^{f} 6 = 6f \quad \text{hence} \quad e \geq 3f.$$

Also, in any polyhedron we have

$$2e = \sum_{i=1}^{v} v_i \geq \sum_{i=1}^{v} 3 = 3v \quad \text{hence} \quad 2e \geq 3v.$$

On the other hand, by using Euler's polyhedral formula we get

$6 = (3f - e) + (3v - 2e)$

which combined with the above two inequalities leads to a contradiction. Hence at least one face should have at most five edges.

A similar solution was submitted by Joshua Lam. The problem was also solved by Rati Gelashvili.

Math Challenges Problems Published in Issue 12 (Fall 2008) of Pi in the Sky

1

Prove that the equation:

$$x^5 + y^5 + 2 = (x + 1)^5 + (y + 2)^5$$

does not have integral solutions

Solution by Edward T. H. Wang

If x and y are of the same parity, then the left side of the equation is even while the right side is odd. If x and y are of opposite parity, then the left side is odd while the right side is even.

Also solved by Rati Gelashvili, and Joshua Lam

2

Prove that $p_n > 2n$ for every $n \geq 5$ where p_n denotes the n^{th} prime number (p_1=2)

Solution by Rati Gelashvili

Let's prove by using induction. The assumption holds for $n = 5$. If $p_{n-1} > 2(n-1)$ then $p_{n-1} \geq 2n - 1$ hence $p_n > p_{n-1} \geq 2n-1$. Since $p_n \neq 2n$ we must have $p_n > 2n$

Also solved by Edward T. H. Wang, Joshua Lam and Andrew J. Pai

3 Inside a square with side length 1 there are 201 points. Prove that there exists a circle of radius 0.1 which contains at least three of these points.

Solution by Rati Gelashvili

The square can be partitioned in 100 little squares of side lengths 0.1. By the Pigeonole Principle at least one square contains at least three points. Since a square of side length 0.1 can be completely covered by a circle with radius 0.1 we conclude that there exists a circle of radius 0.1 which contains at least 3 of the given points.

Also solved by Edward T. H. Wang, Joshua Lam and Andrew J. Pai.

5 Let $ABCD$ be a convex quadrilateral, M the midpoint ofBC and N the midpoint of CD. If $AM + AN = 1$ then prove that the area of the quadrilateral is less than $1/2$.

Solution

We have

$$\begin{aligned} S_{ABCD} &= S_{ABC} + S_{ADC} = 2S_{AMC} + 2S_{ANC} \\ &= 2S_{AMCN} = 2S_{AMN} + 2S_{CMN}. \end{aligned}$$

Since dist(C, MN) < dist(A, MN) (the quadrilateral is convex) we include that $S_{CMN} < S_{AMN}$. Hence $S_{ABCD} < 4S_{AMN}$.

On the other hand

$$\begin{aligned} S_{AMN} &= \frac{1}{2}AM \cdot AN \sin\widehat{MAN} \leq \frac{1}{2}AM \cdot AN \\ &= \frac{1}{2}AM\,(1 - AM) \leq \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}. \end{aligned}$$

Therefore

$$S_{ABCD} < 4\frac{1}{8} = \frac{1}{2}.$$

Also solved by Joshua Lam.

6 There are given five line segments having the property that any three of them can be the sides of a triangle. Prove that at least one of these triangles must be acute.

Solution

Let $a \leq b \leq c \leq d \leq e$ be lengths of the segments. Assume by contradiction that all the constructed triangles are right or obtuse.

Then

$e^2 \geq d^2 + c^2$, $d^2 \geq c^2 + b^2$, $c^2 \geq b^2 + a^2$

and therefore

$e^2 \geq d^2 + c^2 \geq c^2 + b^2 + b^2 + a^2 \geq b^2 + 2ba + a^2$

From this inequality we obtain that $e \geq a + b$, a contradiction.

4 Let a, b, c be positive numbers such that $abc = 18$. Prove that

$$\frac{a^3 + b^3 + c^3}{3} \geq a\sqrt{b+c} + b\sqrt{a+c} + c\sqrt{a+b} \quad (*)$$

Solution by Rati Gelashvili

Since $abc = 18$ then $\sqrt{abc} = 3\sqrt{2}$. If we multiply the left side of the inequality by $3\sqrt{2}$ and the right side by $\sqrt{abc}$ then we need the inequality:
$a^3 + b^3 = (a+b)(a^2 - ab + b^2) \geq (a+b)ab$

(using the fact that $(a - b)^2 \geq 0$ for any a and b).

Using this inequality we have

$$\frac{a^3+b^3}{2\sqrt{2}} + \frac{c^3}{\sqrt{2}} \geq \frac{(a+b)ab}{2\sqrt{2}} + \frac{c^3}{\sqrt{2}} \geq$$

$$2\sqrt{\frac{(a+b)ab}{2\sqrt{2}}\frac{c^3}{\sqrt{2}}} = c\sqrt{abc}\sqrt{a+b}.$$

Notice that we used the inequality $x + y \geq 2\sqrt{xy}$ which is valid for any non-negative numbers x and y.

Similarly

$$\frac{c^3+b^3}{2\sqrt{2}} + \frac{a^3}{\sqrt{2}} \geq a\sqrt{abc}\sqrt{c+b}$$

$$\frac{a^3+c^3}{2\sqrt{2}} + \frac{b^3}{\sqrt{2}} \geq b\sqrt{abc}\sqrt{a+c}$$

Adding these inequalities together, we get

$$\frac{4(a^3+b^3+c^3)}{2\sqrt{2}} \geq \sqrt{abc}(a\sqrt{b+c} + b\sqrt{a+c} + c\sqrt{a+b}).$$

Dividing through by $\sqrt{abc} = 3\sqrt{2}$ and rearranging gives the inequality (*)

Also solved by Joshua Lam using techniques that would not be standard for a general audience.

Math Challenges

Problem 1

Find all positive integers n such that

$$\log_{2008} n = \log_{2009} n + \log_{2010} n.$$

Problem 2

Find the smallest value of the positive integer n such that $(x^2 + y^2 + z^2)^2 \leq n(x^4 + y^4 + z^4)$ for any real numbers x, y, z.

Problem 3

Let a be a positive real number.
Find $f(a) = \max_{x \in \mathbb{R}} \{a + \sin x, a + \cos x\}$.

Problem 4

Prove that the equation $x^2 - x + 1 = p(x + y)$ where p is a prime number, has integral solutions (x, y) for infinitely many values of p.

Problem 5

Find all functions $f : \mathbb{Z} \longrightarrow \mathbb{Z}$ such that $3f(n) - 2f(n + 1) = n - 1$, for every $n \in \mathbb{Z}$. (Here $\mathbb{Z}$ denotes the set of all integers).

Problem 6

In ΔABC, we have $AB = AC$ and $\widehat{BAC} = 100^o$.
Let D be on the extended line through A and C such that C is between A and D and $AD = BC$. Find $\widehat{DBC}$.

.