Never apologize, always explain: Scenes from mathematical life

Peter J. Cameron
G. C. Steward Visiting Fellow
Gonville & Caius College
Cambridge

May 2008

The aim of this short series of lectures is to give some impression of what life as a mathematician is like. It has taken me to many parts of the world, and enable me to work with extraordinary people from many different cultures.

There will be some mathematical content in the lectures. As Julian Havel said, mathematics is not a spectator sport; I will expect some engagement from you.

You may also amuse yourself by spotting several appearances by fellows of Caius in the lectures.

1 Before and beyond Sudoku

In this lecture I am not going to tell you how to solve a Sudoku puzzle. I want to show you two things. First, mathematicians and (more importantly) statisticians had invented all the ingredients of Sudoku some time before the puzzle first appeared. Then I will turn to a variant on Sudoku invented by Robert Connelly and independently by Vaughan Jones, and show you that finding all the solutions can be accomplished by arguments using some of the highlights of finite geometry and coding theory.

But first I am going to discuss the following:

> There's no mathematics involved. Use logic and reasoning to solve the puzzle.
>
> Instructions in *The Independent*

Any mathematician is rightly outraged by this. I would state, very firmly, that mathematics *is* reasoning and logic. The reason for its power and applicability is that the techniques for analysing problems and reasoning your way towards a solution which you learn in a mathematics course generalise to virtually any type of problem which you will meet, and form the best possible equipment for the problem-solver.

As Ian Stewart put it in his book *Does God play dice? The mathematics of chaos*,

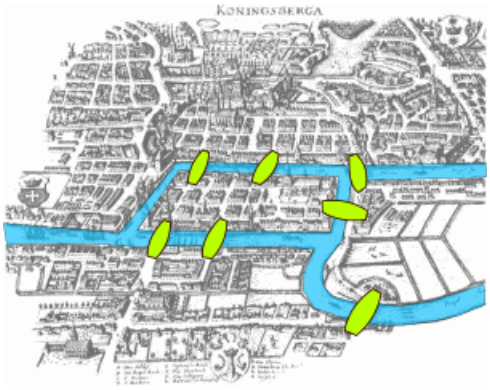
> To criticize mathematics for its abstraction is to miss the point entirely. Abstraction is what makes mathematics work. If you concentrate too closely on too limited an application of a mathematical idea, you rob the mathematician of his most important tools: analogy, generality, and simplicity. Mathematics is the ultimate in technology transfer.

Other mathematicians agree. Sir Michael Atiyah was asked by a journalist what he thought of the current craze for Sudoku. He said that he was pleased to see so many people exercising themselves with mathematical problems every day, and was taken to task on exactly the grounds that *The Independent* would support: unlike Killer Sudoku or Kakuro, Sudoku is not mathematics since you don't actually do *arithmetic* with the numbers (so that any symbols could be used in their place).

1.1 Euler

In a newspaper article on “Ten things you didn’t know about Switzerland” I read, “7. Euler invented Sudoku.” Actually he didn’t, but what he did was very important for later developments. But before I talk about what he did, I should acknowledge that even our hero has feet of clay.

One of Euler’s achievements was the solution of the problem of the *bridges of Königsberg*. Here is the layout of the rivers and bridges in the city of Königsberg (now Kaliningrad) in Euler’ day.



Is it possible to walk around the town, crossing each bridge exactly once? (It is claimed that the citizens of the town liked to walk, annd were very much engaged with this question.) Euler showed that the answer is “No”. The reason is not hard to see. Each region has an odd number of bridges leaving it, so a walk using each bridge exactly once must either start in that region and end somewhere else, or start elsewhere and end in the region. But this is not possible with more than two regions!

Indeed, as is well-known, Euler gave a necessary and sufficient condition for any such network of regions and bridges to have what is now called an *Euler tour*.

But what is less well-known is the contents of a letter he wrote to Carl Ehler, mayor of Danzig, 3 April 1736:

> Thus you see, most noble Sir, how this type of solution [to the Königsberg bridge problem] bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution is based on reason alone, and its discovery does not depend on any mathematical principle …
>
> In the meantime, most noble Sir, you have assigned this question to the geometry of position, but I am ignorant as to what this new discipline involves, and as to which types of problem Leibniz and Wolff expected to see expressed in this way.

It appears, then, that not only did Euler think that reasoning and logic are not the same as mathematics (in fact, he seems to suggest they have little connection), but he admits that he does not know what constitutes the subject of Topology (as we now call "geometry of position") and doesn't seem to think that is mathematics either! What has happened?

My guess is that the word "mathematics" has broadened its meaning since Euler's time.

A recent paper in the *BSHM Bulletin* traces the history of the Königsberg bridges (which have suffered a lot of destruction and rebuilding), indicating in which periods an Euler tour of the bridges was possible. It is now, and the authors celebrated by taking the tour!

1.2 Euler and magic squares

Euler wrote a couple of papers on magic squares. These had been a topic of great interest to mathematicians of many cultures (Chinese and Arabic especially) before they came to Euler's attention. Here is Dürer's *Melancholia*.

16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

In the picture you see a magic square, not very legible, which I have printed to the side for convenience. The entries are the numbers from 1 to 16. Every row, column, or diagonal has sum 34. In addition, we see the date of the engraving, 1514, displayed in the bottom row.

Mathematicians found many constructions for magic squares. Euler wanted to find a construction which would work for all values of n. His idea can be expressed in modern terminology as follows.

A *Graeco-Latin square* of order n is an $n \times n$ array in which each cell contains one of the first n Latin letters and one of the first n Greek letters, in such a way that the following conditions hold:

- any row or column of the array contains each Latin letter once and each Greek letter once;
- given any combination of a Latin and a Greek letter, there is exactly one cell in the array in which that combination occurs.

Of course, other sets of symbols could be used (and must be, if $n > 24$); but the term "Graeco-Latin square" is used whatever the symbol set. This concept was indeed invented by Euler.

Now take the symbols to be $0, \ldots, n-1$ (the digits to base n). Write a two-digit number in each cell: first the entry in A, then that in B. Each two-digit number in base n (from 0 to n^2-1) occurs once, and it is easily checked that all row and column sums are $n(n^2-1)/2$. Add one to each number; then they run from 1 to n^2 and the line sums are $n(n^2+1)/2$. With a bit of fiddling, the diagonal sums can be arranged also. Here's a little example:

$C\beta$	$A\gamma$	$B\alpha$
$A\alpha$	$B\beta$	$C\gamma$
$B\gamma$	$C\alpha$	$A\beta$

21	02	10
00	11	22
12	20	01

8	3	4
1	5	9
6	7	2

So the question is: for which n can a Graeco-Latin square be constructed? Euler knew how to do this for all numbers n not congruent to 2 mod 4, and guessed that it was impossible for the remaining values. This is simple to see in the case $n = 2$ (try it yourself!). The case $n = 6$ was formulated by Euler as follows:

> *Six different regiments have six officers, each one holding a different rank (of six different ranks altogether). Can these* 36 *officers be arranged in a square formation so that each row and column contains one officer of each rank and one from each regiment?*

Trial and error suggests that the answer is "no":

This was not proved until 1900, by Tarry, who exhaustively considered all possible cases. It was not until the 1960s that it was shown that for all larger numbers of this form ($n = 10, 14, \ldots$) there does exist a Graeco-Latin square: Euler was wrong! The three mathematicians who showed this (Bose, Shrikhande and Parker) were referred to as the "Euler spoilers".

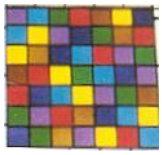
1.3 Statistics

A *Latin square* is an $n \times n$ array in which each cell contains one of the numbers $1, \ldots, n$, in such a way that each number occurs exactly once in each row and once in each column. Of course, the entries needn't be the numbers $1, \ldots, n$, but can be chosen from any set of n symbols, for example, the first n Latin letters (if $n \leq 26$). The term "Latin square" is actually a back-formation from "Graeco-Latin square":

the Latin letters in a Graeco-Latin square form a Latin square (as indeed do the Greek letters).

The question of existence is easy for Latin squares. Place the numbers $1, \ldots, n$ in order in the first row, and then cycle them one place to the right (moving the last one back to the first place) to obtain subsequent rows. However, determining how many Latin squares there are, even approximately, is very difficult and essentially unsolved.

R. A. Fisher pioneered the use of Latin squares in experimental design. He is commemorated in Caius hall by a Latin square:



Anthony Edwards has pointed out to me that there is a mystery about this square, which was taken from the cover of Fisher's book: what is special about it seems to be that nothing is special about it!

Suppose you are conducting an experiment in a field divided into square plots. If there are n different treatments to be applied, and there are $n \times n$ plots, it is sensible to arrange the plots as a Latin square; then, if there is a systematic variation in fertility or soil structure from one side of the field to the other, the treatments will be equally affected. Here is an experiment at Rothamsted experimental station, designed by my colleague Rosemary Bailey. (It has the additional feature that any two treatments occur on neighbouring plots, horizontally or vertically, just once in each order.)

What if there is, for example, a boggy patch in the middle of the field? We should use each treatment once in this patch, as in the following example. This is called a *gerechte design*. (These were invented by W. Behrens; the German word translates as "just", or, more appropriately, "fair".)

1	2	3	4	5
4	5	1	2	3
2	3	4	5	1
5	1	2	3	4
3	4	5	1	2

The final ingredient was provided by John Nelder. It is possible to generate new Latin squares from old by taking a set of positions and moving their entries around to keep the same sets in each row or column; this is called a *trade*. To investigate trades, Nelder defined a *critical set* to be a partially filled Latin square which meets every trade (so that it has a unique completion to a Latin square), but if any entry is removed, there is more than one completion. Here is an example.

1	2		
2			
			3

We now have all the ingredients of Sudoku – a critical set in a gerechte design for the 9×9 square divided into 3×3 subsquares – but it was nearly 20 years later that Harold Garns, a retired architect in New York, invented the puzzle he called "Number Place". It was popularised in Japan by Maki Kaji, who renamed it *Su Doku*. New Zealander Wayne Gould popularised it in the West. The rest is history...

1.4 Symmetric Sudoku

Robert Connelly, a mathematician who works on the stability of Buckminster Fuller-type structures (with elements in tension as well as compression), suggested to me the following variant of Sudoku. (Later and independently Vaughan Jones proposed the same idea.) Suppose that we make the requirements more stringent: each number from 1 to 9 should occur once in each set of the following types:

- rows;
- columns;
- 3×3 subsquares;
- broken rows (one of these consists of three "short rows" in the same position in the three subsquares in a large column);
- broken columns (similarly defined);
- locations (a location consists of the nine cells in a given position, e.g. middle of bottom row, in each of the nine subsquares).

For yourentertainment, here is a symmetric Sudoku puzzle devised by Connelly. You are encouraged to try it before reading further.

								7
				7				
		6						
		4		3				
			1	5				8
					2		7	
					1	4		
					4			
1								

Connelly had given a complete analysis of the solutions to these constraints (which he called "symmetric Sudoku solutions"). His solution can be translated very nicely into finite geometry, as I will now describe.

First we coordinatise the Sudoku grid. Let F denote the finite field with three elements $\{0, 1, 2\}$ (the integers mod 3). We label each cell with an element of F^4 as follows:

- the first coordinate is the number of the large row containing the cell;
- the second coordinate is the number of the row within the large row;
- the third coordinate is the number of the large column;
- the fourth coordinate is the number of the column within the large column.

We really have an *affine space* rather than a vector space. (An affine space is essentially a vector space with the special role of the origin removed. So affine

subspaces are cosets of vector subspaces. For example, the unique affine line through the points a and b has the form $\{(1-t)a+tb : t \in F\}$. Now the elements of F are $0,1,2$; substituting these values for t shows that the three points on the line are $a, b, -a-b$. In other words, *three points form a line if and only if they add up to zero*. Moreover, a set of points in an affine space is an affine subspace if and only if it contains the line through any two of its points.

Now the regions defining symmetric Sudoku are cosets of the 2-dimensional affine subspaces with equations as follows:

- rows: $x_1 = x_2 = 0$;
- columns: $x_3 = x_4 = 0$;
- subsquares: $x_1 = x_3 = 0$;
- broken rows: $x_2 = x_3 = 0$;
- broken columns: $x_1 = x_4 = 0$;
- locations: $x_3 = x_4 = 0$.

So the positions of any given symbol must differ in at least three coordinates. This means that they form a 1-*error-correcting code*. Here is a small digression on coding theory.

We have to send messages through a noisy channel; each message is a word or string of symbols in a fixed alphabet, and there is a small probability that one symbol will be changed into another during transmission. We arrange to send messages that look sufficiently different from one another that if a small number of errors are made the original message can still be recognised.

For example, suppose that all codewords differ in at least three positions. If one symbol is changed during transmission, the received word is only one step away from the transmitted word, but at least two steps from any other word. So the transmitted word can be recognised.

In our case, each cell of the Sudoku grid is represented by a 4-tuple of elements of F, and as we noted above, the set of positions of a fixed symbol has the property that any two of the 4-tuples differ in at least three places: thus, a 1-error correcting code, as claimed.

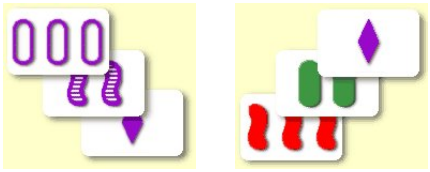
Counting shows that this code is "perfect", that is, any 4-tuple differs in at most one position from exactly one codeword. The rather surprising geometric interpretation of this is that the balls of radius 1 with centres at the codewords cover the entire space F^4 without any overlapping.

It is known that there is just one type of perfect 1-error correcting code of length 4 over F, the so-called *Hamming code*. So a symmetric Sudoku solution is precisely the same as a partition of F^4 into nine Hamming codes!

Our next goal is to show that the Hamming codes actually form affine subspaces. As we noted, a subspace is characterised by the property that it contains the line joining any two of its points. We also noted that three points form a line if and only if they add up to zero.

In the field F (the integers mod 3), a moment's thought shows that three elements a,b,c satisfy $a+b+c=0$ if and only if either a,b,c are all equal, or they are all different: e.g. $2+2+2=0$ and $0+1+2=0$. So three points form a line if and only if, in each of their four coordinates, the entries are either all equal or all different.

This can be nicely understood in terms of the card game SET, some of whose cards are as shown here:

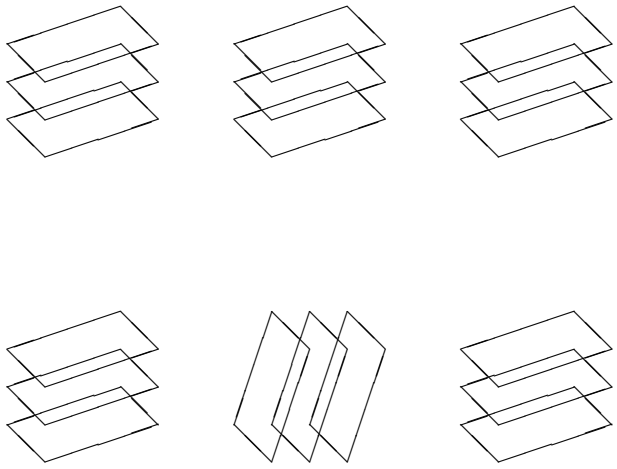


Each card has four attributes (number, colour, shape and filling), each of which can take three values. So the set of cards is matched with F^4. A "SET" (a winning combination in the game) consists of three cards such that, for each attribute, either they are all the same, or they are all different. The examples given are both SETs. (If you are reading this in black-and-white, the symbols on the three cards on the left are all purple; those on the right, from the top down, are purple, green, and red.)

The SETs are thus precisely the lines of the affine space. So we have to show that a perfect code has the property that, given any two of its elements, the third point forming a SET with them is also in the code. Strangely (or maybe not so strangely), this is a typical Sudoku argument! We know the positions of two occurrences of a given symbol, and infer the position of a third occurrence, just as you do solving Sudoku puzzles; in a sense, it is easier, since we have more constraints. (If you solved the puzzle given earlier, you probably discovered this principle for yourself.)

Once we have reached this conclusion, it is easy to show that there are just two essentially different ways to partition F^4 into nine Hamming codes, and hence two symmetric Sudoku solutions. For one of them, we simply take one Hamming

code (forming a plane in the space) and all the planes parallel to it. For the other, we switch the three planes in a 3-dimensional space into a different set of three parallel planes. This schematic picture shows how. You see that, if two sets of parallel planes are switched, we could have started from a different parallel class and switch one set, so no new solutions are obtained.



The problem of constructing a symmetric Sudoku solution was subsequently posed as a problem in *Emissary*, the newsletter of the Mathematical Sciences Research Institute in Berkeley, by the Fields Medallist Vaughan Jones.

Our account of all this and much else appeared in the May 2008 issue of the *American Mathematical Monthly* (we in this context being Rosemary, Bob, and I).

A final remark of interest to Sudoku players. A Sudoku puzzle is a set of entries in a gerechte design containing a critial set; so we would like to know the size of the smallest critical set (the smallest number of entries in a valid Sudoku).

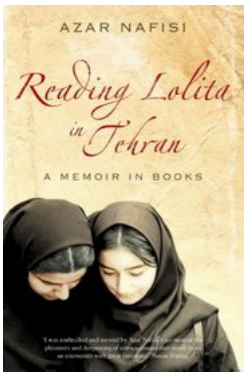
For Latin squares, the number is conjectured to be a quarter of the total number of cells, rounded up: that is, $\lceil n^2/4 \rceil$. For $n = 9$, this would give 21. However, the conjecture has only been proved for $n \leq 8$.

For Sudoku, the extra costraint should make the number smaller. It is widely conjectured that the answer is 17. Many 17-entry puzzles are known, and extensive computer searches by Gordon Royle have failed to find one with fewer than 17 entries.

Of course, the same square can have critical sets of different sizes, and for different squares, the sizes of the smallest critical sets can be different.

2 Proving theorems in Tehran

(after *Reading Lolita in Tehran: A Memoir in Books* by Azar Nafisi)



In 2003, an international conference on Combinatorics, Linear Algebra and Graph Colouring was held at the Institute for Studies in Theoretical Physics and Mathematics (IPM) in Tehran, Iran. A select group of international visitors was invited, and the process of obtaining an Iranian visa was expedited for us. (The Rough Guide suggests that Iranian visas take weeks or months to obtain, but I received mine by return post from the Iranian Embassy in London.)

Because of teaching commitments, I had to take an overnight flight which arrived on the morning the conference began. No problem: I was met at the airport by a car and driver from the IPM, driven through the truly appalling traffic, and taken to the guesthouse, where I had just enough time to have breakfast of bread and honey and tea before we left for the Institute. Mine was the first lecture of the conference: a surge of adrenaline got me through the talk.

Before the conference, an unsavoury incident had occurred. All invited speakers had received emailed death threats warning us not to come to Iran or we would be killed. The threats were sent several times in increasingly strident language. Almost without exception, we were not deterred. But we soon noticed that there were two strong silent men with bulging briefcases on the bus with us for every journey. Clearly the authorities were taking no chances.

Also on the bus with us was Mandana, our hostess, about whom I will say more later.

The conference was one of the most memorable I have ever attended. Every day there was a daily bulletin giving some interesting information: a recipe for *bagali polo* which we had for lunch[1]; the history and contents of the museums to which we were taken in the many free afternoons; the mathematical genealogy of the invited speakers; competitions for the students. (One of the competitions was to discover the middle names of the invited speakers. The winner solved the problem in the most straightforward way, by asking us! The prize was to have her photograph taken with us.)

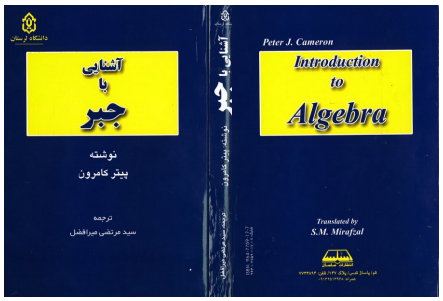
Most spectacular of all was the conference excursion, a two-day trip to the ancient city of Isfahan, a town with a river spanned by many ancient bridges,

[1] See appendix

a ruined Zoroastrian fire temple, many gorgeous palaces and mosques, and the second largest town square in the world (smaller than Tiananman Square, but larger than Red Square.)

I had decided to stay for a few days after the conference, in the hope that there might be a chance of doing some mathematics with locals or other visitors. So indeed it turned out. In the journal volume containing the conference papers, as well as my own paper and my edition of the problems from the problem session, I had no less than three joint papers containing results found in those few days after the conference.

As well as this, I had the chance to get to know Mandana better. When, a couple of years later, I came to read Azar Nafisi's book *Reading Lolita in Tehran*, I saw some similarities between her story and Mandana's. Mandana was a student of English literature, who would have liked to make a living by translating books into Farsi. But typically she would produce a translation and apply to have publication approved, only to be told "Why do you waste your time translating that decadent rubbish?" As a result, she supported herself teaching English to the children of Korean migrant workers, and odd jobs like acting as hostess to a group of visiting mathematicians. She also wrote poetry, in both Farsi and English. She volunteered to translate some of my poems into Farsi, but warned me I wouldn't get any money for this, since Iran does not subscribe to the copyright convention. (I knew this already. The translator of my algebra textbook, S. M. Mirafzal, introduced himself to me and gave me a copy of the Farsi edition, whose existence I hadn't even dreamed of.)



I now want to describe two pieces of work I did during the conference or my stay in Tehran afterwards. A third piece of work will appear briefly in the next lecture.

2.1 Self-dual, not self-polar

An *incidence structure* consists of a set of "points" and a set of "blocks", with a relation of "incidence" between points and blocks. Many people identify a block with the set of points incident with it, and regard an incidence structure as a set of "points" with a distinguished collection of subsets called "blocks". But this approach, by giving a special role to the points, destroys the symmetry between points and blocks inherent in the first definition I gave.

For example, in real plane projective geometry, we can take any theorem and form its dual, by interchanging the labels "point" and "line". The resulting statement will make sense; more, it will be true, since the real projective plane is *self-dual*, in the sense defined below.

Let (P,B,I) be an incidence structure. (This means that P is the set of points, B the set of blocks, and I the incidence relation between them.) A *duality* consists of a pair of bijective functions $f : P \to B$ and $g : B \to P$ which "reverse incidence", in the sense that if point p and block b are incident, then so are the point $g(b)$ and the block $f(p)$. An incidence structure is *self-dual* if a duality exists; this means that the structure is isomorphic to its *dual* (with the labels "point" and "block" reversed).

It is natural to wonder when we can take g to be the inverse of f. We say that the duality is a *polarity* if this holds. There are several ways to express this. If we apply a duality twice, the resulting maps take P to P and B to B and preserve incidence; that is, they comprise an automorphism of the structure. A duality is a polarity if and only if its square is the identity automorphism. Another say of describing a polarity, just in terms of the map f, is that for any two points p and q, p is incident with $f(q)$ if and only if q is incident with $f(p)$. Yet again, a structure is *self-polar* if and only if the incidence is represented by a symmetric matrix.

Willem Haemers remarked that rather elaborate constructions, using properties of finite simple groups of Lie type, show the existence of incidence structures (of a special type called *generalized quadrangles*) which are self-dual but not self-polar. The smallest such example has 85 points, and fairly elaborate computations in vector spaces over finite fields are required to verify the property. Haemers asked for a simpler example. We were able to come up with one with just eight points. On my return to London, my colleague Donald Preece was able to reduce this to seven points. Moreover, we could show that seven is the smallest possible number of points in such an example, and that there are just four different incidence structures with seven points which are self-dual but not self-polar.

The strategy is as follows. An incidence structure can be represented by a

bipartite graph, whose vertices are the points and blocks, an edge joining two vertices if they are incident. (This is called the *incidence graph* or *Levi graph* of the structure. Levi was the teacher of Bose, one of the "Euler spoilers" we met in the last chapter.) Now an automorphism of the incidence structure is a graph automorphism which fixes the two bipartite sets (points and blocks), while a duality is a graph automorphism which interchanges them. So we are looking for a bipartite graph in which the two bipartite sets can be interchanged by an automorphism of order 4, but not by one of order 2. I leave it as an exercise for you to find one with 14 vertices.

2.2 Symmetric sign patterns

Another of the invited speakers at the meeting, Charles Johnson, was interested in the following question: What do I learn about the eigenvalues of a symmetric real matrix if I know only the signs of its entries? Call a matrix a *symmetric sign pattern* if it has all entries $+1$ or -1 and is symmetric. An arbitrary real matrix M with all entries nonzero *conforms* to a sign pattern P if the sign (positive or negative) of the entry m_{ij} is p_{ij} for all i, j. (We are considering only matrices with no zero entries.) Johnson's question is: what properties of symmetric matrices are forced by the sign pattern?

The operation of *switching* a symmetric $n \times n$ matrix (or sign pattern) with respect to a subset J of $\{1,\ldots,n\}$ is performed by changing the sign of the rows and columns with index in J. The switched matrix is similar to the unswitched one, so interesting properties are preserved. Similarly, simultaneous permutation of rows and columns gives a similar matrix. Also, changing the sign of all entries just negates the matrix and does not essentially change its properties. So Johnson wanted to know: How many symmetric sign patterns are there, up to switching, permutations, and negation? He told me the numbers for $n = 1,2,3,4$: they were 1, 2, 4, 11 respectively.

As it happened, these numbers were very familiar to me: they are the numbers of graphs on n vertices (up to isomorphism) for $n = 1,\ldots,4$. This suggested an obvious conjecture.

In fact, it is clear that the conjecture is true for odd n. For in this case, any symmetric sign pattern is equivalent (under switching and negation) to a unique pattern with all row and column sums even; the off-diagonal -1 entries in this pattern determine the adjacency in a graph, But for even n, there is no natural bijection between graphs and symmetric sign patterns, so more is required.

I was well prepared for this. A graph is *even* if every vertex lies on an even

number of edges. (In terms of Euler's work on the Königsberg bridges, a graph has a closed Euler tour if and only if it is connected and even.) In the early 1970s, Mallows and Sloane had given a direct proof that the numbers of even graphs on n vertices is equal to the number of *switching classes* of n-vertex graphs (these can be regarded as equivalence classes of symmetric sign patterns with zero diagonal under switching and permutation). They proved this by finding a formula for the number of switching classes and showing that it agrees with the known formula for even graphs. Once again, there is a simple reason for this if n is odd, since each switching class contains a unique even graph. I had found a more conceptual proof of the general statement, using the notion of duality in vector spaces. Combining the two approaches led us to the proof of the conjectured answer to Johnson's question.

In this case we didn't need it, but there is a very valuable Internet resource for investigations of this kind: Neil Sloane's *On-Line Encyclopedia of Integer Sequences*. If you find an unknown integer sequence, in almost any kind of investigation, look it up in the Encyclopedia: chances are someone has found it before, probably in a different context, and you will learn something from the connection. (In fact, it was while compiling data for the first, printed version of the Encyclopedia, that Mallows and Sloane noticed the equality of the numbers of switching classes of graphs and even graphs, and set about proving it.) I will say more about the Encyclopedia in the last lecture.

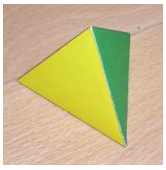
A third topic I worked on in Tehran will make a brief appearance in the next lecture. I will outline it here.

For reasons connected with the theory of designs, which I won't elaborate on, we wanted to find out all possible sizes for a collection of k-element subsets of the set $\{1,\ldots,n\}$ fixed by the action of a certain permutation group (the group known as $\mathrm{PSL}(2,q)$, where $n = q+1$). This required three things

- knowledge of all the subgroups of $\mathrm{PSL}(2,q)$ (this is "classical", having been worked out by Dickson in the early 20th century);
- knowledge of their orbit sizes (this was worked out not long after);
- knowledge of the so-called "Möbius function" of each subgroup (this is also known, though more recent and more obscure).

There are three "strange" subgroups here which don't form part of a general pattern. It turns out that they are the groups of rotations of the regular polyhedra (tetrahedron, cube amd dodecahedron). (Note: although there are five regular

polyhedra, there are only three different groups, since each of two dual pairs – the cube and octahedron, and the dodecahedron and icosahedron, have the same group.)



These groups will reappear in the next chapter.

Appendix

Lima Bean with Dill Rice BAGALI SHEVID POLOW
Bagali polo
(Serves 6 to 8)

The following recipes are from 'Secrets of Cooking' by Linda Chirinian (ISBN 0-9617033-0-X Lionhart Inc. New Canaan, CT).
This exotic Iranian dish can be served with plain yogurt spooned over the rice, or with roast chicken, barbecued lamb chops, or steak. A straight-sided, non-stick, saucepan is the best kind of pot to use for this recipe.

Ingredients:

- PREPARATION TIME: 20 MINUTES (plus soaking for rice)
- COOKING TIME: 45 MINUTES

- 1 recipe Steamed Rice
- 1 package (10 ounce) frozen baby lima beans, thawed
- 14 tablespoons butter
- 3 cups freshly chopped dill
- 3 medium potaties, cut into 1 inch slices (optional)
- 1/4 teaspoon cinnamon threads crushed and steeped in 2 tablespoons hot water
- salt and freshly ground pepper to taste
  - Prepare steamed rice. Melt 4 tablespoons butter in a non-stick 6-quart saucepan. Arrange potato slices in single layer in saucepan.

  - Spread one-third of prepared rice over potatoes. Salt and pepper. Cover with half of lima beans, and half of dill. Cover with half of remaining rice and remainder of lima beans and dill. Top with remaining rice. Keep ingredients mounded high in center so steam can circulate. Sprinkle 4 cups water over rice. Slice remaining butter, place over rice. Cover rice with waxed paper.
  - Cook over medium-high heat 8 minutes, reduce heat to low, and cook 35 minutes or until rice is soft and fluffy.
  - Set 1 cup rice aside. Mound remaining rice on serving dish. Remove potatoes from saucepan with spatula and place around rice or in separate dish. Sprinkle reserved cup of rice with saffron and mix well. Spread saffron rice on top of plain rice. Season with salt and pepper.

- VARIATION: When layering rice, add 1 large onion, chopped and sauteed in butter, 6 broiled lamb chops or 2 pounds cooked boneless lamb shoulder cubes, or 6 cooked chicken cutlets. Increase cooking time by 15 minutes.

3 Transgressing the boundaries

(with apologies to Alan Sokal, "Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity", *Social Text*, Spring/Summer 1996)

There are many ways of dividing up mathematicians, some jovial ("There are 10 types of mathematician, those who understand binary notation and those who don't"), some serious. Tim Gowers wrote an essay on "The two cultures of mathematics", the title based on C. P. Snow's "two cultures" (artistic and scientific). Roughly put, Gowers distinguished the type of mathematics which requires the building of big theories from that which solves problems. The theory-builders look down on the problem-solvers because they do not prove impressive theorems; in fact what they do is to develop techniques which may be applicable to a wide variety of problems.

I suppose I stand more on the side of the problem-solvers. The prototypical subject of the theory-builders is algebraic geometry. I have never proved a theorem of algebraic geometry, but I once used a little light algebraic geometry (dimension theory) to prove a result about the growth function of the number of orbits of an infinite permutation group on n-element sets of the domain.

This brings me to my own division of mathematicians, into those who delve deep and those who range more widely. Here I am definitely in the second camp. I work mainly in algebra and combinatorics, but have had papers published in the *Annals of Pure and Applied Logic*, *Probability Theory and Related Fields*, and the *Journal of Mathematical Psychology*, among others. The most satisfying results of mine are those in which an idea from one field has turned out to be useful in another, quite different, area.

I want to discuss three cases of this from my own work. But first a brief description of an example due to Sokal himself.

3.1 Chromatic roots

A *graph* consists of a set of vertices, some pairs of which are joined by edges. If every pair is joined, it is a *complete graph*; if no pairs are joined, it is a *null graph*. The complete graph on n vertices is denoted by K_n.

A *proper colouring* of a graph G with q colours is an assignment of the colours to the vertices of G in such a way that two adjacent vertices get different colours. For any graph, there will be a certain minimum number of colours needed to do

this: for example, for K_n, all the vertices must be coloured differently, so we need at least n colours. It can be shown that there is a polynomial P_G associated with a graph G, so that the number of colourings with q colours is $P_G(q)$ for any natural number q.

The famous *four-colour conjecture*, proved by Appel and Haken in 1977, asserts that if a graph G can be drawn in the plane without edges crossing, then it can be coloured with four colours; in other words, 4 is not a root of the equation $P_G(q) = 0$. This led to a lot of interest in the roots (not necessarily integers) of chromatic polynomials. These are called *chromatic roots*. It was shown that there are no negative chromatic roots, none between 0 and 1, and (remarkably) none between 1 and $32/27$; but after $32/27$, chromatic roots are dense (that is, there is are chromatic roots arbitrarily close to any real number).

Interest turned to complex chromatic roots. Part of the incentive came from physics, in particular the study of *phase transitions* (e.g. ice melting, iron becoming magnetised). These are apparently discontinuous "macroscopic" changes in materials in which all the "microscopic" quantities behave continuously. This is the subject matter of *statistical mechanics*, Sokal's main field.

The Nobel prizewinners Lee and Yang had postulated a mechanism for phase transitions which involved complex roots of certain polynomials approaching the real line arbitrarily closely. The polynomials were generalisations of chromatic polynomials of graphs.

Partly in view of the fact that there are no negative chromatic roots, and partly based on computations with reasonably small graphs, it was conjectured that complex chromatic roots cannot have negative real parts. However, this was blown out of the water when Sokal showed that actually complex chromatic roots are dense in the whole complex plane. (So, although there are no negative real roots, complex roots come arbitrarily close to the negative real axis).

I won't describe Sokal's elegant proof here, but the driving principle behind it comes naturally to a physicist. Think of a graph as an electrical network. You probably imagine each edge as a 1 ohm resistor. However it is better to let the resistance of each edge be a parameter which can take arbitrary values, even complex ones. (A circiut including inductance and capacitance as well as resistance can be thought of as having complex resistance.) Then we can use the series and parallel formulae from electrical circuit theory to simplify our graph, replacing several edges in series or parallel with a single edge. This enabled Sokal to find general results for a particular class of graphs called theta-graphs. At the end of the calculation, he simply had to put all the resistances equal to 1 to obtain the graph-theoretic conclusion.

3.2 Automata, permutation groups, and cores

An *isomorphism* from a graph G to a graph H is a bijective mapping between the vertex sets that carries edges to edges and non-edges to non-edges. A weaker notion is that of a *homomorphism*, a mapping that takes edges to edges – we do not care what it does to a non-edge, which may be mapped to a non-edge, or to an edge, or collapsed to a single vertex. The symbol $G \to H$ means that there is a homomorphism from G to H.

Two graphs are called *hom-equivalent* if there are homomorphisms in both directions between them. This is an equivalence relation on the class of all finite graphs. It is known that there is a unique (up to isomorphism) smallest graph hom-equivalent to any given graph, called its *core*. The core of a graph G is realised as an induced subgraph of G.

Here is an example. The *clique number* $\omega(G)$ of a graph G is the largest m such that G contains a complete graph on m vertices: that is, the largest m for which $K_m \to G$. As we saw earlier, the *chromatic number* $\chi(G)$ is the smallest number of colours required to colour the vertices so that adjacent vertices receive different colours; that is, the smallest m such that $G \to K_m$. (Take a moment to see why this is true. Graph homomorphisms generalise graph colouring and are widely used to formulate constraint satisfaction problems.)

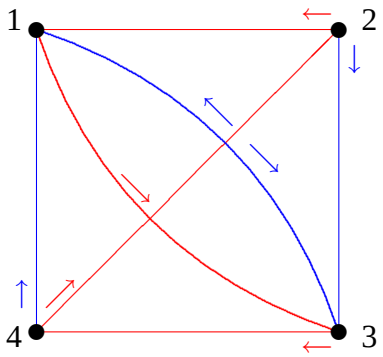
Clearly, for any graph G we have $\omega(G) \leq \chi(G)$; equality holds if and only if the core of G is a complete graph.

It is known that the core of a graph inherits some symmetry properties of the graph, such as vertex-transitivity. When I was asked by a former postdoc, Cristy Kazanidis, for a research problem to work on, I suggested looking at the cores of graphs with a great deal of symmetry. We chose to look at the so-called *rank* 3 *graphs*, those whose automorphism groups act transitively on vertices, edges and non-edges. After looking at a few cases, we came up with the conjecture that either the core of such a graph is complete, or that the graph is itself a core.

The next chapter of the story comes from a completely different source, indirectly from two researchers in automata theory, João Araújo in Lisbon and Ben Steinberg in Ottawa. Following notes of João, I can describe the problem this way. Imagine you are in a dungeon consisting of a number of interconnecting rooms. Passages between rooms are marked with coloured arrows. Each room contains a special door; in one room, the door leads to freedom, but in all the others, to instant death. You have a schematic map of the dungeon, but you do not know where you are located. You would like to have a sequence of colours such that, following the edges of these colours in order from any starting point, you will ar-

rive at the escape room. Such a sequence is called a *reset word* in automata theory. Not every finite automaton has a reset word; but the oldest problem in automata theory, the *Černý conjecture*, states that, if a reset word exists, then there is one of length at most $(n-1)^2$, where n is the number of states (or rooms in our example).

Here is an example, due to João Araújo.



You can check that (Blue, Red, Blue, Blue) is a reset word which takes you to room 3 no matter where you start.

A related question, the *road-colouring problem*, has been solved recently. Given a directed graph with a constant number d of edges out of each vertex, can it be coloured with d colours so that there is a reset word? The obvious necessary conditions are that the graph is *strongly connected* (there is a directed path from any vertex to any other) and the lengths of the directed cycles are coprime. These were conjectured to be sufficient by Benjamin Weiss and Roy Adler in 1970; this conjecture was proved last year by Avraham Trahtman.

If every colour corresponds to a permutation of the states, then no reset word exists. The two researchers were led to the concept of a *synchronizing* permutation group, one with the property that if any non-permutation is added to it, it is possible to generate a reset word (that is, a constant function).

I learned about this by two different routes: first, from a lecture in summer 2007, which described a seemingly different concept for permutation groups and stated that it arose from automata theory; and in January this year, when a former PhD student of mine, now in Ottawa, told me about a conversation he had had with Ben Steinberg at a bus stop. Very soon I was able to see that this was connected with cores. Specifically, I was able to prove the following theorems:

- If the automorphism group of a graph G acts transitively on the non-edges of G, then either the core of G is a complete graph, or G is itself a core. (Thus the conjecture Cristy and I made is true.)

- A permutation group is non-synchronizing if and only if it is contained in the automorphism group of a non-null graph which is not a core.

3.3 Root systems and line graphs

A graph can be specified by giving its *adjacency matrix*. If the vertices of G are $v_1, \ldots, v_n$, the adjacency matrix $A(G)$ is the $n \times n$ matrix with (i, j) entry 1 if v_i is joined to v_j, and 0 otherwise. The algebraic properties of $A(G)$, in particular its eigenvalues, give important information about the graph. Note that writing the vertices in a different order replaces A by $P^{-1}AP$ for some permutation matrix P, and doesn't change the eigenvalues. Also, we may assume that the graph is connected; if not, it is enough to look at the connected components.

In the 1950s and 1960s there was a lot of interest in the smallest eigenvalue $\lambda(G)$ of a graph G. It is easily shown that $\lambda(G) \leq -1$, with equality if and only if G is a complete graph (assuming it is connected). Attention turned to graphs with least eigenvalue -2 or greater. After several special cases had been analysed, Alan Hoffman wrote a long manuscript dealing with the general case.

Part of the difficulty comes from the fact that there are two families of graphs with this property: line graphs and cocktail party graphs. Let H be a graph. The *line graph* of H is the graph $L(H)$ whose vertices are the edges of H, two vertices of $L(H)$ being joined if and only if the corresponding edges have a common vertex. Let N be the vertex-edge incidence matrix of H. Then $N^\top N = 2I + A(G)$, where $G = L(H)$. Since $N^\top N$ is positive semi-definite (all eigenvalues non-negative), $A(G)$ has all eigenvalues -2 or greater.

A *cocktail party graph* has $2m$ vertices $v_1, \ldots, v_m$, $w_1, \ldots, w_m$; all pairs are joined except v_i and w_i for all i. (At a cocktail party, you talk to everybody except your own partner.) It is an easy exercise to show that this graph has smallest eigenvalue -2.

Hoffman discovered a class of graphs which he called *generalized line graphs* also fitting the bill. One of these is obtained by taking the line graph of a graph H, and glueing on cocktail parties to the edges through each vertex of H.

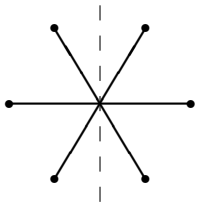
Hoffman's manuscript, which was not published (I believe he was not completely happy with the argument) purported to prove that all "sufficiently large" graphs with least eigenvalue -2 or greater is a generalized line graph. In the meantime, Jean-Marie Goethals in Brussels, Jaap Seidel in Eindhoven, Ernie Shult in Manhattan, Kansas, and I found a much nicer proof of the theorem which showed exactly how large "sufficiently large" has to be.

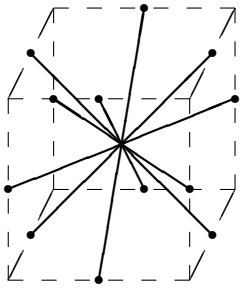
The proof uses the notion of a root system, developed by Cartan and Killing in order to classify the simple Lie algebras over the complex numbers. A *root system* is a finite set S of non-zero vectors in real Euclidean space with the following properties:

- if $v, cv \in S$ then $c = \pm 1$;
- if $v, w \in S$ then $2(v \cdot w)/(v \cdot v)$ is an integer;
- S is mapped to itself by the reflection in the hyperplane perpendicular to any of its vectors. (The reflection corresponding to v is the map $w \mapsto w - 2(v \cdot w)/(v \cdot v)v$.)

A root system is *indecomposable* if it is not contained in the union of two non-zero orthogonal subspaces.

It is impossible to draw helpful pictures of high-dimensional objects; but here are root systems in 2 and 3 dimensions, known as A_2 and A_3 respectively. They are beautiful, symmetric objects; Mark Ronan, in his recent book on symmetry, refers to them as multidimensional "crystals".





The part of the classification we need is that of indecomposable *spherical* root systems (where all roots have the same length). There are two infinite families of these, called A_n and D_n, where n is a positive integer, and three "exceptional" ones, E_6, E_7 and E_8.

I visited Eindhoven regularly at the time. (Their budget went by calendar years and had to be spent before the end of the year, so I often visited Eindhoven in December.) When I arrived, Jaap told me of his discovery: first, any graph with least eigenvalue -2 or greater can be represented as a set of Euclidean vectors. For $2I + A$ is positive semi-definite; and any positive semi-definite symmetric matrix is the matrix of inner products of a set of Euclidean vectors.

Jaap had also discovered that the maximal such sets formed an infinite family with three exceptions.

The next day we drove to Brussels to visit Jean-Marie. He explained to Jaap that he had also found one infinite family and three exceptions. But on comparison, we found that they had different infinite families. Now two infinite families and three exceptions meant one thing to me: spherical root systems. And so it turned out.

The trick is very simple. The vectors representing the graph satisfy $v_i \cdot v_i = 2$ and $v_i \cdot v_j \in \{0,1\}$ for $i \neq j$. So the lines they span are at angles 90° or 60°. Call three lines in a plane mutually at angles 60° a *star*. A small calculation (this is the only calculation required in the proof) shows that, if a set of lines at angles 90° and 60° contains two lines of a star, then the third line of the star also makes angles 90° or 60° with all the lines. Thus, a maximal set is *star-closed*. If we take vectors of fixed length in both directions along all the lines, the first two conditions in the definition of a root system are clearly satisfied, and star-closure gives the third condition.

Everything was simple after that. Although A_n is maximal in n dimensions, it is contained in D_{n+1} (one dimension higher), so we could ignore A_n. It turned out that graphs embeddable in D_n are precisely Hoffman's generalized line graphs! Of course, only a finite number of graphs can be embedded in an exceptional root system. (The largest, E_8, has 240 vectors, but an embedded graph has at most 36 vertices.)

The resulting paper is one of my most-cited, and has found various applications. But so far nobody has managed to classify the graphs with least eigenvalue -3 or greater: this is a much harder problem!

3.4 Connection number and Möbius function

The third topic is an open problem, a guess at a connection based on very slim evidence. Before stating it, I mention briefly another case involving one of the protagonists in this story, John McKay.

The largest sporadic simple group is the so-called "Monster", a group of order

$$808\,017\,424\,794\,512\,875\,886\,459\,904\,961\,710\,757\,005\,754\,368\,000\,000\,000.$$

Before this group was actually constructed, it was shown (on the basis of a hypothesis later confirmed) that the smallest size of complex matrices representing the group is 196883.

In the classical nineteenth-century theory of complex functions, the *modular function* plays a particular role. It is a function defined on the upper half-plane

and invariant under the group of Möbius transformations $z \mapsto (az+b)/(cz+d)$, where a,b,c,d are integers with $ad-bc=1$. The Laurent series of the modular function begins

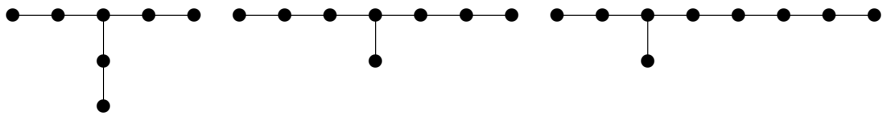
$$z^{-1} + 196884z + \cdots$$

McKay asked: Is it just coincidence that $196884 = 196883 + 1$? Of course, there were not too many mathematicians in the world who would have been aware of the significance of both these numbers at the time: finite group theorists do not study complex analysis or *vice versa*!

The answer is that it is not at all coincidence, and McKay's observation led to the theory of "monstrous moonshine" (the term was invented by John Conway), with connections to Lie algebras and conformal field theory, and eventually to a Fields medal for Conway's student Richard Borcherds.

Now to another place where McKay played a role. There is a strange and shadowy parallelism between, on the one hand, the three "dual pairs" of regular polyhedra (the tetrahedron, cube/octahedron, and dodecahedron/icosahedron), and the three exceptional root systems E_6, E_7, E_8 mentioned in the last section. I will briefly discuss this before turning to the open problem.

Each rotation group G in 3-dimensional space can be "lifted" to a group $\overline{G}$ twice as large consisting of 2×2 complex unitary matrices containing $-I$; the factor group $\overline{G}/\langle -I \rangle$ is isomorphic to G. From the groups of the regular polyhedra (the tetrahedral, octahedral and icosahedral groups), we get corresponding *binary* groups of orders 24, 48 and 120. Now we consider the irreducible complex matrix representations of this group; they are the vertices of a graph in which two representations R_1 and R_2 are joined by an edge if and only if R_2 is a constituent of $R_1 \otimes S$, where S is the distinguished representation of degree 2. The graphs in the case of the three groups are shown below. (The fact that S is unitary implies that edges are undirected.)



Now take a spherical root system. The set of integer linear combinations of roots forms a *lattice* in the Euclidean space, a discrete additive subgroup. There is a basis of so-called "fundamental roots", such that each root is a linear combination of fundamental roots with all coefficients of the same sign. We can add a special root to the fundamental set in a canonical way which I don't want to explain here. Now the inner product of two roots from the enlarged set is 0 or -1:

form a graph by joining two roots if their inner product is -1. (If A is the adjacency matrix of the graph, then the inner product matrix is $2I - A$; so the greatest eigenvalue of A is at most 2. Compare this with what we had in the last section. Now it turns out that the "extended Coxeter–Dynkin diagrams" (as they are called) for the exceptional root systems E_6, E_7, E_8 are precisely the three graphs above. (This is the *McKay correspondence*. It works for the root systems A_n and D_n as well, suitably re-interpreted – there are no regular polyhedra for these!)

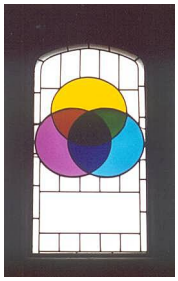
Put another way, the extended Coxeter–Dynkin diagrams are precisely the connected graphs whose vertices can be labelled with numbers in such a way that the sum of the labels of the neighbours of any vertex is twice the label of the vertex. Exercise: find such labellings for the graphs in the above figure.

Other areas of mathematics including singularity theory also show the same correspondence between regular polyhedra and exceptional root systems, or more generally, provide occurrences of the root systems and their Coxeter–Dynkin diagrams. Indeed, these diagrams are among the most ubiquitous objects in the grammar of mathematics. In the 1970s, the American Mathematical Society asked a select group of mathematicians to propose problems to replace the famous "Hilbert Problems" from 1900, which influenced the course of mathematics in the twentieth century. The Russian mathematician V. I. Arnol'd proposed the problem of explaining the ubiquity of the Coxeter–Dynkin diagrams.

Last year, some of my colleagues and ex-students arranged a conference for my sixtieth birthday, which was held during four days of perfect weather in Ambleside, in the Lake District. The conspirators, recalling that I had once said that any advanced alien civilisation would certainly know the Coxeter–Dynkin diagrams, arranged for my son to incorporate them in a birthday card.

Now to the problem. In Tehran, I worked with three postdocs at the IPM (Maimani, Omidi and Tayfeh-Rezaie), on counting the designs admitting certain groups. We had to be able to count subsets of the domain invariant under certain subgroups of the group in question (this is easy), and from this, to count subsets invariant under the given subgroup but no larger one. The tool for this is a generalisation of the Inclusion–Exclusion Principle called *Möbius inversion*.

In ordinary Inclusion–Exclusion, we have a set of properties and we know how many elements satisfy any given subset of the properties; we can compute the number which satisfy that subset and no others. The result is obtained by adding up the numbers satisfying all larger sets of properties with appropriate coefficients, which turn out to be all $+1$ and -1. In terms of a Venn diagram, if we know how many elements are in the whole box, in each of the circles, and in all the intersections of circles, we can count the number of elements excluded by all the circles.



In the generalisation, we do a similar calculation with subgroups in place of sets of properties, but the coefficients are not necessarily $+1$ and -1. The *Möbius function* is the function $\mu(H,K)$ defined on pairs of subgroups H,K with $H \leq K$: if $a(H)$ is the number of sets fixed by H, then the number fixed by H and no more is $\sum_{K \geq H} \mu(H,K)a(K)$.

The crucial value is $\mu(1,G)$ where 1 is the identity and G the whole group. It turned out that for the groups we were looking at, the hardest subgroups to deal with were exactly the polyhedral groups. The value of $|\mu(1,G)|$, where G is the tetrahedral, octahedral, or icosahedral group, is 3, 2, or 1 respectively. (There is a sign, which I will ignore here.)

Now it also happens that if $L(R)$ is the root lattice of the root system R (the set of integer linear combinations of R) and $L^*(R)$ is the dual lattice (the set of vectors v such that $v \cdot r$ is an integer for all $r \in R$), then $L^*(R)/L(R)$ is a finite abelian group; its order is called the *connection number* of the root system. For $R = E_6, E_7, E_8$, the connection numbers are 3, 2, 1 respectively.

Is there a connection??

4 Cameron felt like counting

> 'I count a lot of things that there's no need to count,' Cameron said. 'Just because that's the way I am. But I count all the things that need to be counted.'
>
> Richard Brautigan, *The Hawkline Monster*

Like the character in Brautigan's novel, I have always enjoyed counting, from a very early age. I will talk about some counting problems, solved and unsolved.

Ernst Mach said, "There is no problem in all mathematics that cannot be solved by direct counting." This may overstate the case, but counting is absolutely fundamental to mathematics.

First, we should make clear that counting can mean many different things. Let us suppose that we are counting "objects" of some kind with "size" n. Let $f(n)$ be the number of such objects.

- Find a formula for $f(n)$.
- Find a *recurrence relation* which allows $f(n)$ to be calculated from knowledge of the preceding values $f(0), \ldots, f(n-1)$.
- Find explicitly the *generating function* $\sum_{n=0}^{\infty} f(n)x^n$ – then, at least in principle, the coefficients can be found by analytic methods.
- Give a (hopefully efficient) algorithm for calculating $f(n)$: ideally quicker than generating the objects and counting them!
- Failing an exact formula, give upper and lower bounds for $f(n)$, as close together as possible.
- Find an *asymptotic formula* for $f(n)$, that is, a (hopefully simple) function $g(n)$ such that $f(n)/g(n) \to 1$ as $n \to \infty$.

There are several other things we might want, closely related to counting. Among these might be

- An algorithm to generate all the objects, or to move from one object to the next in some suitable ordering.
- A method to choose an object at random (all objects equally likely), or failing that, a method which makes the probabilities approximately equal, and closer to equal the longer we are prepared to spend on it.

4.1 Partitions and representations

The result of counting objects can have interest beyond the count itself. Here is a very famous example, concerning the symmetric group S_n (the group of all permutations of $\{1,\ldots,n\}$).

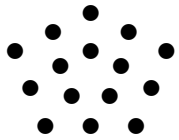
A *partition* of n is an sequence of positive integers with sum n, arranged in non-increasing order. For example, there are five partitions of 4:

$$4 = 3+1 = 2+2 = 2+1+1 = 1+1+1+1.$$

To make a small digression, counting partitions is a very famous problem, to which Euler made the first serious contribution. No simple formula is known for $p(n)$, but it satisfies the recurrence relation

$$p(n) = p(n-1)+p(n-2)-p(n-5)-p(n-7)+p(n-12)+p(n-15)+\cdots,$$

where the sequence continues as long as the numbers inside the brackets are non-negative. The terms have the form $(-1)^{k-1}p(n-k(3k\pm1)/2)$; the numbers $k(3k\pm1)/2$ are *pentagonal numbers*, since they record the number of dots in a pentagonal arrangement of side k; so the recurrence is called *Euler's Pentagonal Numbers Theorem*.



Why does this recurrence satisfy our criteria? The number of terms in the recurrence for $p(n)$ is about the square root of n; so we can compute $p(1),\ldots,p(n)$ with approximately $n^{3/2}$ additions and subtractions. This calculation was actually performed by Percy McMahon in the early 20th century, and his data led to the brilliant work of Ramanujan and Hardy on partitions.

Back to the story. First, a bit of theory shows that the irreducible representations of S_n can be "labelled" by partitions n. There is an ordering on the partitions. Now we can let the symmetric group act on r-tuples $(A_1,\ldots,A_r)$ of pairwise disjoint sets with union $\{1,\ldots,n\}$, where $|A_i| = a_i$ for all i. Take the action of the group on this set by permutation matrices, and decompose it. You will find just one irreducible module in the decomposition which has not occurred for any partition earlier on the list. This is the representation indexed by the partition $(a_1,\ldots,a_n)$. What is its degree?

The partition can be represented by n empty boxes placed in r rows with a_i boxes in the ith row, for $i = 1, \ldots, r$, the boxes being aligned on the left. Given such a diagram, a *tableau* is a filling of the boxes with the numbers $1, \ldots, n$ such that the numbers increase along every row and down every column. The number of tableaux for a given partition is precisely the degree of the corresponding irreducible representation!

For example, here are the five tableaux asssociated with the partition $5 = 3+2$; so there is a corresponding representation of degree 5 of S_5.

1	2	3
4	5	

1	2	4
3	5	

1	2	5
3	4	

1	3	4
2	5	

1	3	5
2	4	

4.2 Permutations

A *permutation* of the set $\{1, 2, \ldots, n\}$ is a rearrangement of the elements; in other words, a function mapping the set to itself which is invertible (that is, one-to-one and onto). Any permutation can be represented in *cycle notation*, in which we track each element as it is repeatedly mapped by the permutation until it returns to its starting point. For example, the permutation of $\{1, 2, 3, 4, 5\}$ which maps $1 \to 2$, $2 \to 3$, $3 \to 1$, $4 \to 5$ and $5 \to 4$ would be represented as $(1, 2, 3)(4, 5)$. If the permutation leaves a point where it is, that point forms a cycle of length 1.

There are many interesting counting problems regarding permutations. Often they can be conveniently expressed in terms of generating functions, of a particular type called exponential generating functions. If we have a class containing a_n permutations of $\{1, \ldots, n\}$, we take the *exponential generating function* to be

$$A(x) = \sum_{n=0}^{\infty} \frac{a_n x^n}{n!}.$$

(The name comes from the fact that, for the simplest possible sequence, with all terms equal to 1, we obtain the series for the exponential function.

Now the exponential generating function (or e.g.f. for short) of the sequence counting all permutations is

$$P(x) = \sum_{n=0}^{\infty} \frac{n!\, x^n}{n!} = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

For reasons that will appear, we only want to take the terms with even index; so we get the even powers of x, and our new function is

$$P'(x) = \frac{1}{1-x^2}.$$

In connection with a problem involving Latin squares, I observed that the e.g.f. for permutations in which all cycles have even length is

$$E(x) = \frac{1}{\sqrt{1-x^2}}.$$

(Here we only get even powers of x, since such a permutation can only exist if n is even.) Now there is a multiplicative principle for generating functions: For an arbitrary permutation, there is a unique way of splitting it up into a permutation with all cycles even and one with all cycles odd. It follows that, if $O(x)$ is the e.g.f. for permutations with all cycles odd acting on an even number of points, then

$$E(x) \cdot O(x) = P'(x).$$

From our formulae above, we conclude that

$$O(x) = \frac{1}{\sqrt{1-x^2}} = E(x).$$

In other words,

> *The numbers of permutations with all cycles even and all cycles odd are equal.*

For example, when $n = 4$, there are nine permutations with all cycles even (three products of two 2-cycles, and six 4-cycles), and nine with all cycles odd (the identity, and eight permutations consisting of a 3-cycle and a fixed point).

This cries out for a "direct" proof, that is, one matching up the two types of permutations. I asked for such a proof at the British Combinatorial Conference in 1993. On the way home from the conference, Richard Lewis and Simon Norton independently found such a bijection, and wrote it up as a joint paper in the conference proceedings. You might like to try this one yourself. It is not a very "natural" bijection, though!

4.3 Latin squares

We met the concept of a Latin square in the first lecture, and saw how they occur in various parts of mathematics and statistics. How many Latin squares are there?

This is a very difficult questions, and heroic calculations by Brendan McKay and Ian Wanless have managed to count Latin squares up to order 11: the number of Latin squares of order 11 is

$$776\,966\,836\,171\,770\,144\,107\,444\,346\,734\,230\,682\,311\,065\,600\,000.$$

You may think that we are over-counting, and that there are some obvious equivalences: permuting the rows, the columns, or the symbols in a Latin square gives a different Latin square which is obviously "equivalent" to the first. One can also transpose a Latin square, and there are more subtle operations involving exchanging rows and symbols, or columns and symbols. In this way, the set of all Latin squares is partitioned into *main classes*. But these are harder to count, and the number of main classes has only been calculated up to order 10.

In fact it doesn't matter too much. Clearly the number of Latin squares is at most n^{n^2}, since this is the number of ways of filling an $n \times n$ array if n symbols are available. You can probably see how to improve this bound slightly. But it is known that the number is at least $(n/c)^{n^2}$, which is not too much smaller. Put another way, we know that the *logarithm* of the number of Latin squares is asymptotic to $n^2 \log n$. Now the number of Latin squares in the same main class as a given one cannot exceed $(n!)^3 \cdot 6$ (permutations of rows, columns, and symbols among themselves, and permutations of the three classes: the group of such transformations is known as the *wreath product* of S_n by S_3.) This number is much smaller, asymptotically, than the number of Latin squares, that dividing by it has little effect: the difference this makes is swallowed up in our ignorance about the magnitude of the total number.

Failing this, one of our alternatives was choosing a Latin square at random. There is another motivation for this. One of the first statisticians to use Latin squares, R. A. Fisher, introduced the idea of "randomization" in experimental design, to validate the analysis of variance. Fisher advocated choosing a random Latin square for an experiment based on Latin squares; for this purpose, he and Yates began the tabulation of Latin squares which has been continued by McKay and Wanless. (However, Fisher's method is no longer required; it is enough to choose any Latin square and apply a random element of the wreath product of S_n by S_3 to it.)

There are theoretical reasons why we still need a method for choosing Latin squares at random. Since there are so many of them, we want to know something about their "typical properties". For example, each row of a Latin square is a permutation of $\{1,\dots,n\}$; what is the distribution of the number of rows which are odd permutations? A natural conjecture would be that it is approximately the same as the number of heads in n tosses of a fair coin (a binomial distribution).

A method for choosing random Latin squares was found by Jacobson and Matthews. It involves a random walk through the very large space. But it is not easy to think of a small step to take from one Latin square to another. The big idea of Jacobson and Matthews was to enlarge the space to contain "improper" Latin squares containing one "fault".

Suppose, for example, that the entry in row i and column j is k, and we wish to change it to k'. The element k' already occurs in row i (in position j', say) and in column j (in row i', say), and we should change these entries to k to avoid duplications in these rows. If we are lucky, the entry in row i' and column j' will be k; then changing it to k' will give us a Latin square. But it is much more probable that there is a different entry, say k'', in cell (i',j'). In this case, we change this cell so that it contains two entries k' and k'' and one "negative" entry k. If you then check, you will see that, allowing for signs, each symbol still occurs once in each row and once in each column.

		j		j'	
i		k		k'	
i'		k'		k''	

$\longrightarrow$

		j		j'	
i		k'		k	
i'		k		$k'+k''-k$	

If we start from an improper Latin square with one fault, we must fix the fault: if a cell contains $k'+k''-k$, we must change it to either k' or k'', and make consequential changes elsewhere. This may lead to one fault somewhere else in the square.

If we walk around the space making such changes, we reach genuine Latin squares not too infrequently, and as the number of steps increases the distribution of the Latin squares we reach tends to the uniform distribution.

This method is easy to implement, and has been used to explore questions about typical Latin squares. For example, Thomas Prellberg and I looked at the question of the number of odd rows. It seems that it is close to a binomial distribution, but the tails are a little heavier than they should be; in other words, the

number of Latin squares with all rows of the same parity is slightly more than you would expect if the parities were really random and independent. But we can't prove anything yet!

4.4 Parking

In the 1960s, the following problem was raised (originally in connection with data storage by computers).

> *A car park has n spaces in a line, numbered* $1,2,\ldots,n$*. The drivers of n cars have each independently decided on the position where they want to park. As each driver arrives at the car park, (s)he drive to the preferred parking place. If the space is free, (s)he parks there. If not, (s)he drives on and takes the first available space; if (s)he doesn't find an empty space, (s)he leaves in disgust.*
>
> *What is the probability that all drivers manage to park?*

If $n = 2$, the probability is $\frac{3}{4}$, since only if both drivers choose 2 will they fail to park.

The answer is surprisingly simple, and there is a beautiful argument to show it. The required probability is $\dfrac{(n+1)^{n-1}}{n^n}$. Said otherwise, out of the n^n ways the drivers can make their choices, exactly $(n+1)^{n-1}$ lead to everyone parking successfully. A beautiful argument was found by Henry Pollak of Bell Labs.

Consider instead a circular car park with $n+1$ spaces. Again there are n drivers, and the same rules apply; but now everybody will succeed in finding a place, since each driver continues round the circle until reaching a free spot. Now the parking is successful in the linear car park if and only if space $n+1$ is unoccupied in the circular car park (i.e. nobody chooses $n+1$, and nobody is forced to park there). By symmetry, any space in the circular car park is equally likely to be unoccupied: so, out of the $(n+1)^n$ choices, a fraction $1/(n+1)$ lead to successful parking in the linear car park.

A year or two ago, I wondered about the number k of drivers who are unable to park. This number is a random variable: can its distribution be calculated? I calculated it for some small values, and Emil Vaughan (a Ph.D. student at Queen Mary) pushed it further.

At that point, I put the problem on my web page, so that all the world could have a go. The challenge was taken up by two computer science students from Saarbrücken in Germany, Daniel Johannsen and Pascal Schweitzer. They were

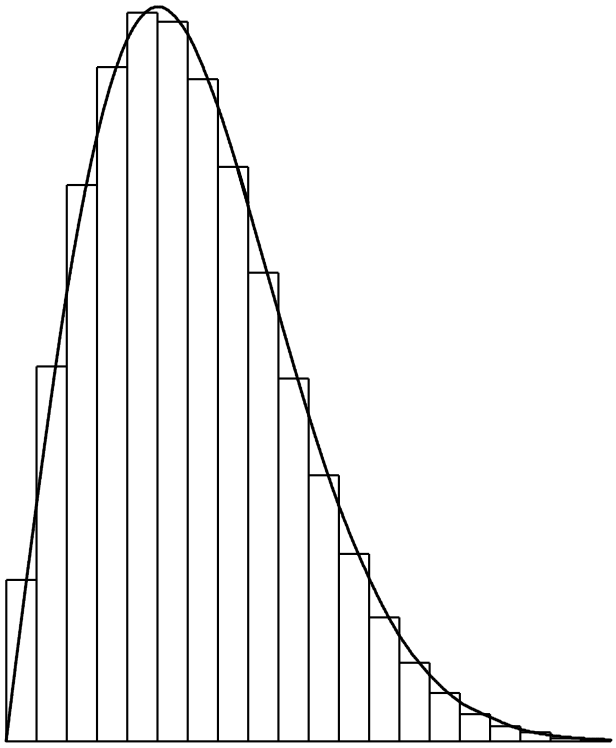
able to find a recurrence relation for the numbers, and even more impressively, to solve it to give an explicit (though rather complicated!) formula.

I wanted to see a plot of the data. For this, the scale is very important. For example, since the limit of $(1+1/n)^n$ as $n \to \infty$ is e (the base of natural logarithms), we see that the probability that $k = 0$ (everybody parks) is about $\mathrm{e}/(n+1)$. If we scale the y-axis by a factor of n, this will be visible, but it turns out that some other values will be much too large to appear on the page.

My colleague Thomas Prellberg took up the challenge. He found that the correct scaling was by the square root of n on the y-axis, and by $1/\sqrt{n}$ on the x-axis. If this is done, moreover, the histogram of the probabilities tends to a limit known as the *Rayleigh distribution*. I have shown a plot of the histogram for $n = 100$ (for $k > 20$, the probabilities are too small to appear on the graph), and on the comparable scale, the p.d.f. of the Rayleigh distribution. The curve has equation $y = 4x\mathrm{e}^{-2x^2}$.

The result is a bit surprising. The Rayleigh distribution arises classically as the length of a random vector in the Euclidean plane whose X and Y coordinates are independent normal variables with mean zero and the same variance. There is no obvious connection with our parking problem!

In calculating the distribution function, Thomas had to evaluate a rather complicated integral. He was so pleased when he found the right substitution to do this that he set it as a prize question for the first-year Calculus class; one student managed to claim the prize.



(This section is based on an article to appear in the forthcoming magazine *QED*.)

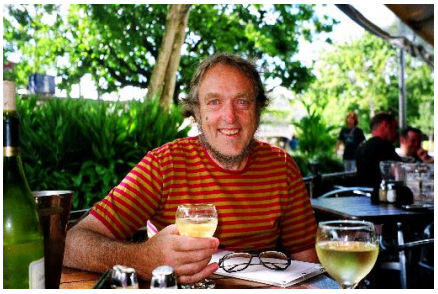
4.5 The On-Line Encyclopedia of Integer Sequences

To anyone who counts things, the On-Line Encyclopedia of Integer Sequences, maintained by Neil Sloane at AT&T Bell Labs, is a resource without price (and also without cost). Look up the sequence you have just found; it is very likely that you will either find it under a completely different name (and so discover that what you are working on is closely related to another branch of mathematics, whose insights you can use), or that it is a transform of another known sequence (in which case you have to figure out why this is).

Let me briefly describe my own first acquaintance with this resource. This dates from the mid 1970s, in the days before on-line resources. I was studying infinite permutation groups, and had just constructed a new interesting example, with a complicated construction involving sequences of rational numbers. I was interested in the number of orbits of the group acting on the set of n-element sub-

sets, and calculated the first few numbers. I knew that Neil Sloane was about to publish a book containing some interesting sequences (about 2000 of them), and that my colleague Dominic Welsh had a preliminary version of the book. I phoned him up and read out the numbers to him: "$1,2,3,6,11,23,46$" "Don't you mean 47?" "No, I am sure it is 46." It turned out that my sequence was not the same as the one in the book, which counted certain kinds of trees; but, by a remarkable coincidence, my sequence counted trees of a different kind. Finding this connection led to some very fruitful research. (For the record, my sequence is number A001190 in the Encyclopedia, labelled "Wedderburn–Etherington numbers", while the one Dominic found is number A000055, "Number of trees".)

The Encyclopedia celebrated its 100K birthday recently (the addition of the 100 000th sequence). Here you see the host, Neil Sloane, inviting contributors to the virtual birthday party, and one of the contributors toasting the Encyclopedia.



Acknowledgements Thanks to Neill Cameron for the picture of Euler's officers, and for the birthday card with the Coxeter–Dynkin diagrams; and to Sue Welham for the picture of the experimental field at Rothamsted Experimental Station; and to my colleagues and collaborators for their contributions to the enjoyment I have had from mathematics! And of course, thanks to the Master and Fellows of Gonville & Caius College for the opportunity to present these lectures, and especially to Jonathan Evans for making the arrangements.