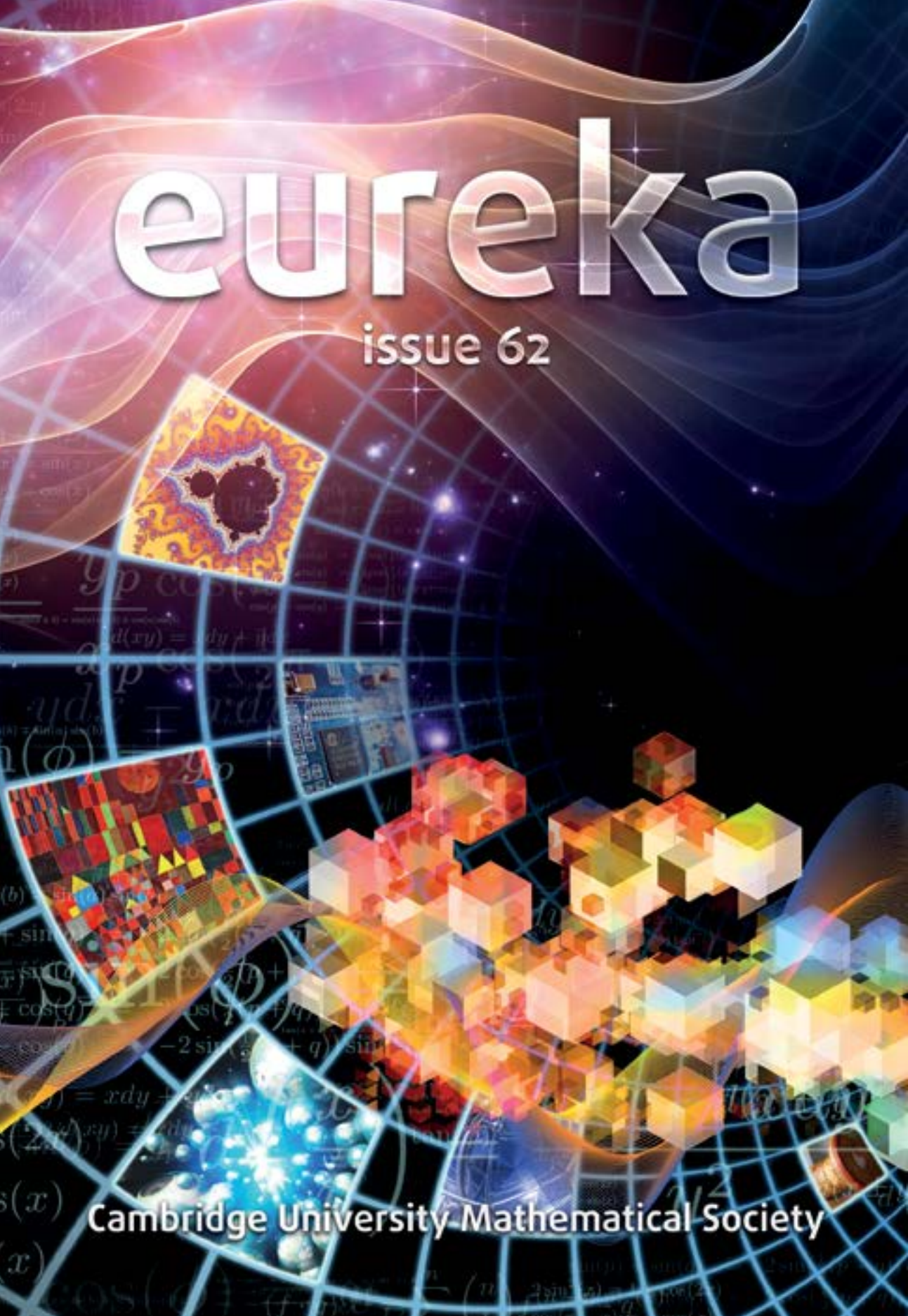
eureka

issue 62



Cambridge University Mathematical Society

Eureka 62

A Journal of The Archimedeans
Cambridge University Mathematical Society

Editors: Philipp Legner and Jack Williams
Cover by Andrew Ostrovsky, Inner Cover by George Hart

December 2012

Editorial

Eureka 62

When the Archimedeans asked me to edit *Eureka* for the third time, I was a bit sceptical. Issue 60 was the first to get a paperback binding and issue 61 was the first to be published in full colour and with a new design. How could we make this issue *special* – not just a repeat of the previous one?

Eureka has always been a magazine for students, not a research journal. Articles should be interesting and entertaining to read, and often they are a stepping stone into particular problems or areas of mathematics which the reader would not usually have encountered.

Every year we receive many great articles by students and mathematicians. Our task as editors is often to make them more visually appealing – and we can do so using images, diagrams, fonts or colours.

What we wanted to add in this issue was interactivity, such as videos, slideshows, animations or games. Unfortunately this still is quite difficult on paper, so we decided to publish a second version of Eureka as interactive eBook for mobile devices like iPad. And we hope that this will make reading mathematics even more engaging and fun.

The digital version will, for the first time, make Eureka available to a large number of students outside Cambridge. And therefore we have reprinted some of the best articles from previous issues. We spent many hours in the library archives, reading old copies of Eureka, though of course there are many more great articles we could have included.

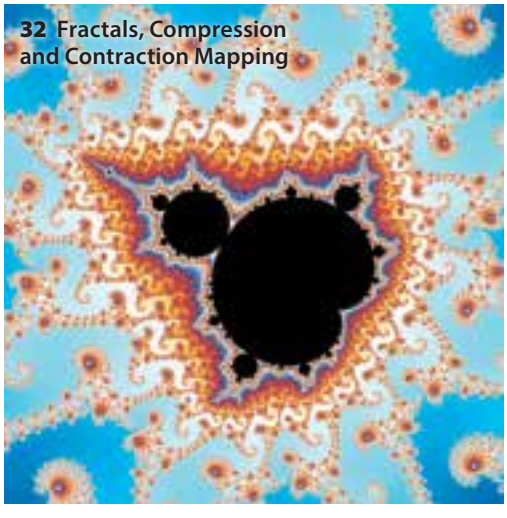
The articles in this issue are on a wide range of topics – from number theory to cosmology, from statistics to geometry. Some are very technical while others are more recreational, but we hope that there is something interesting for everyone.

I want to thank the editorial team for all their work, and the authors for their excellent articles. We hope you will enjoy reading Eureka 62!

Philipp Legner and Jack Williams

Contents

First magic number in nuclear physics.
$\sqrt{2}$ was the first known irrational number.



Items marked † are reprinted from past issues of Eureka.

The Archimedeans

Yuhan Gao, President 2012 – 2013

This year was yet another highly successful one for The Archimedeans. The society welcomed over 150 new members, courtesy of a very popular Freshers' Squash. We hosted a number of talks given by speakers from the university over the course of Michaelmas and Lent. These covered a number of different topics, catering for those with interests in pure, applied and applicable mathematics. Highlights included talks by Prof. Grae Worster on Ice, and Prof. Imre Leader on Games of Pursuit and Evasion.

The society expanded the range of events which we offered to our members this year. We held a board games evening, which proved to be a thoroughly enjoyable night for all those who attended. One of our most anticipated events was the black-tie Annual Dinner in the delightful surroundings of the Crowne Plaza Hotel.

A tradition of the Archimedeans is to hold an annual Problems Drive. This time around, teams from as far afield as Oxford came to take part in an engaging and entertaining mathematics competition. Prizes were awarded not only for the teams with the highest scores, but also for particularly creative team names. The questions given can be found in this journal, and we welcome you to try them yourself.

The year finished on a high in May Week, courtesy of the Science and Engineering Garden Party. Six societies from the university joined together to host a brilliant afternoon of fun, aided by a jazz band. Finger food and Pimm's was served, and there was even a cheese bar on offer.

We would like to thank our members for contributing to an excellent year for the society. I would also like to thank the committee for all of their hard work, and Philipp Kleppmann, last years' President, along with the previous committee, for everything which they have done for the society. We look forward to another exciting year ahead.

The Committee 2012 – 2013

President
Yuhan Gao *(Trinity)*

Vice-Presidents
Sean Moss *(Trinity)*
Dana Ma *(Newnham)*

Corporate Officer
Joseph Briggs *(Trinity)*

Secretary
Jacquie Hu *(Jesus)*

Treasurer
Colin Egan *(Gonville and Caius)*

Events Managers
Pawel Rzemieniecki *(Fitzwilliam)*
Yuming Mei *(Emmanuel)*

Publicity Officer
James Bell *(Gonville and Caius)*

Webmaster
Ben Millwood *(Downing)*

Smallest composite number. Number of Nucleobase types in the DNA: A, G, C and T.

▲ Professor David Tong

THE AR

Archimedeans ▶ Garden Party

◀ Archimedeans Problems Drive

CHI

Archimedeans ▼ Annual Dinner

▲ Archimedeans Talk in the CMS

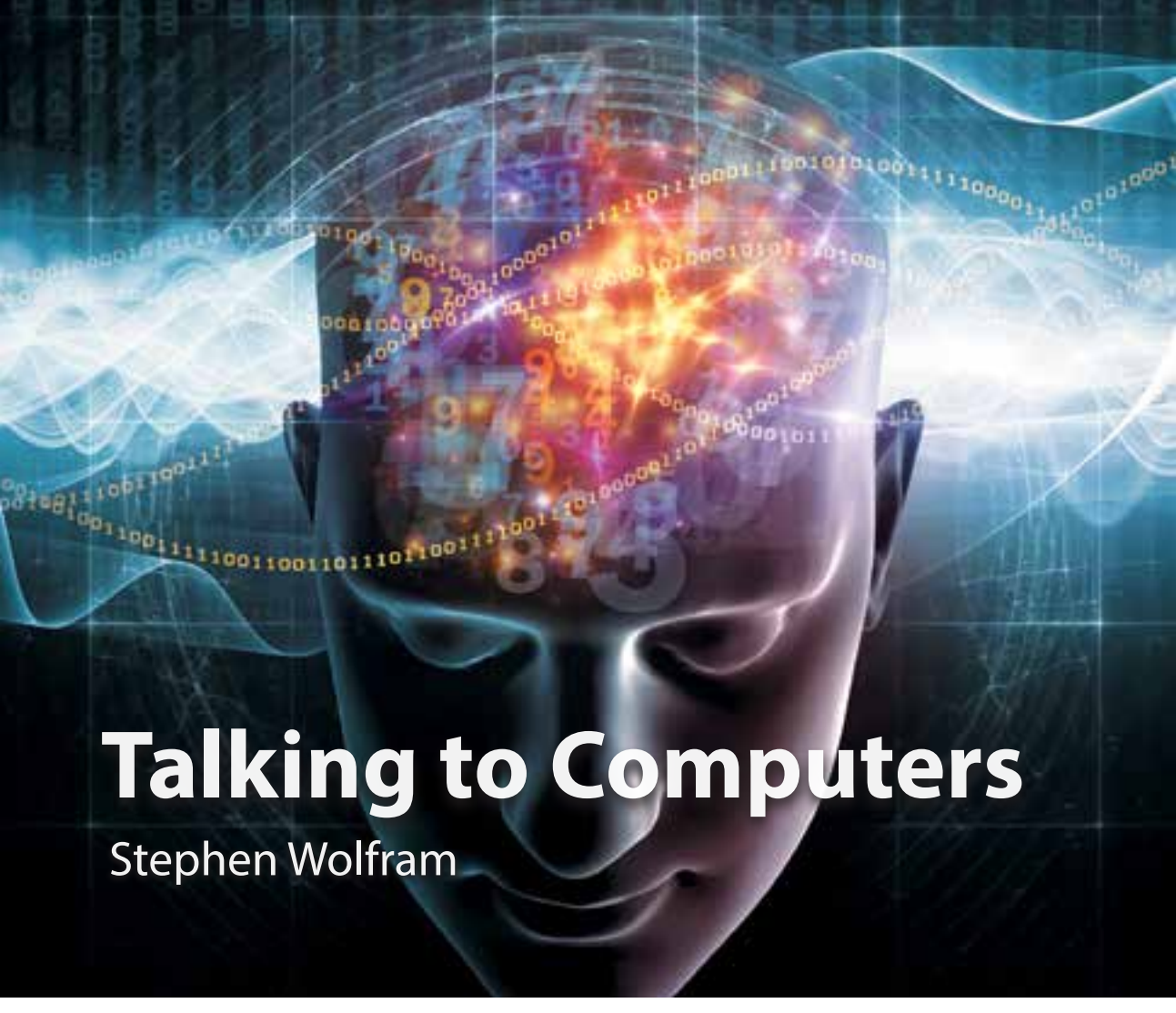
MEDE

◀ Archimedeans Problems Drive

◀ Archimedeans Garden Party

ANS

K_5 is the smallest complete non-planar graph.
The cycle of fifths underlies musical harmonies.



Talking to Computers

Stephen Wolfram

I love computer languages. In fact, I've spent roughly half my life nurturing one particular very rich computer language: Mathematica.

But do we really need computer languages to tell our computers what to do? Why can't we just use natural human languages, like English, instead?

If you had asked me a few years ago, I would have said it was hopeless. That perhaps one could make toy examples, but that ultimately natural language just wouldn't be up to the task of creating useful programs.

But then along came Wolfram|Alpha in which we've been able to make free-form linguistics work vastly better than I ever thought possible.

But still, in Wolfram|Alpha the input is essentially just set up to request knowledge – and Wolfram|Alpha responds by computing and presenting whatever knowledge is requested. But programming is different. It is not about generating static knowledge, but about generating programs that can take a range of inputs, and dynamically perform operations.

The first question is: how might we represent these programs? In principle we could use pretty much any programming language. But to make things practical, particularly at the beginning, we need a programming language with a couple of key characteristics.

The most important is that programs a user might specify with short pieces of natural language must typically be short – and readable – in the computer language. Because otherwise the user won't be able to tell – at least not easily – whether

Smallest perfect number. Hexagonal tilings give the densest 'sphere' packing in two dimensions.

the program that's been produced actually does what they want.

A second, somewhat related criterion is that it must be possible for arbitrary program fragments to stand alone – so that large programs can realistically be built up incrementally, much like a description in natural language is built up incrementally with sentences and the like.

To get the first of these characteristics requires a very high-level language, in which there are already many constructs already built in to the language – and well enough designed that they all fit together without messy "glue" code.

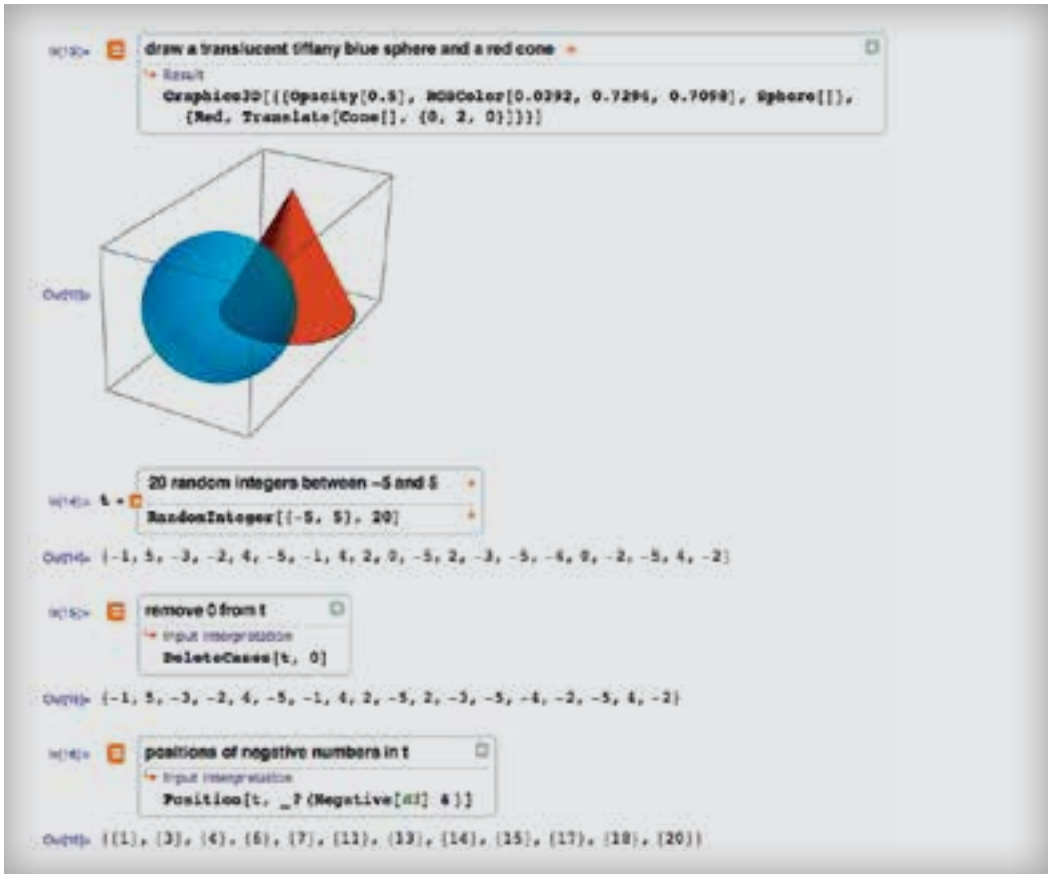
And to get the second characteristic essentially requires a symbolic language, in which any piece of any program is always a meaningful symbolic expression.

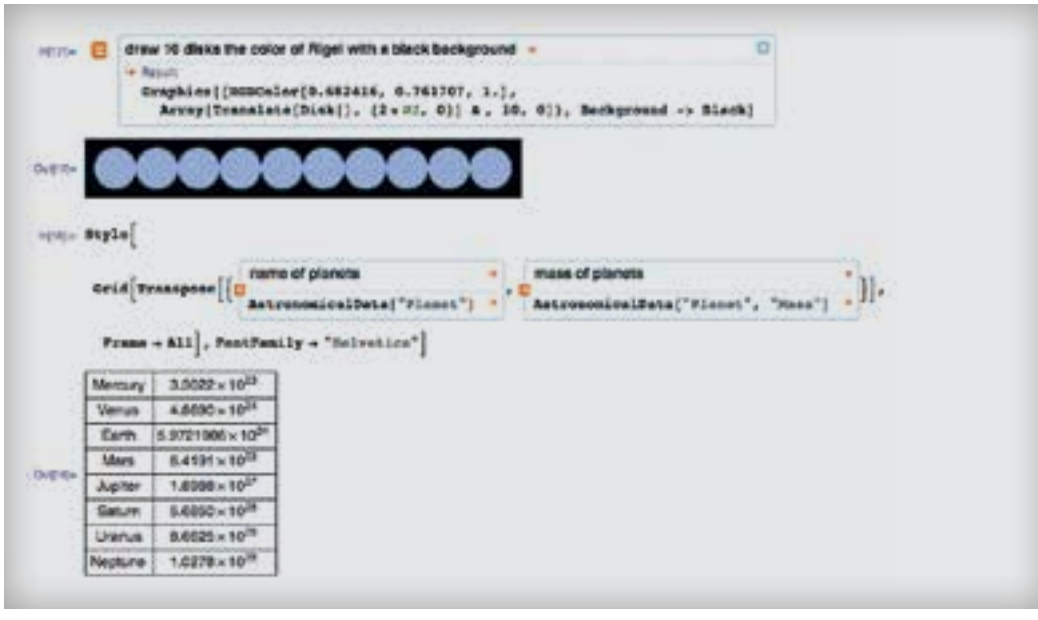
Conveniently enough, there is one language that satisfies rather well both these requirements: Mathematica!

The linguistic capabilities of Wolfram|Alpha give one the idea that one might be able to understand free-form natural language specifications of programs. Mathematica is what gives one the idea that there might be a reasonable target for programs generated automatically from natural language.

For me, there was also a third motivating idea – that came from my work on *A New Kind of Science*. One might have thought that to perform any kind of complex task would always require a complex program. But what I learned in *A New Kind of Science* is that simple programs can often do highly complex things.

And the result of this is that it's often possible to find useful programs just by searching for them in the computational universe of possible programs – a technique that we use with increasing frequency in the actual development of both Wolfram|Alpha and Mathematica.





And it was this that made me think that – even if all else failed – one might be able to "synthesize" programs from natural language just by searching for them.

OK – so there are reasons to hope that it might be possible to use natural language input to do programming. But can one actually make it work?

Even when Wolfram|Alpha was launched, I still wasn't sure. But as we worked on bringing Wolfram|Alpha together with Mathematica, I got more and more optimistic.

And with Mathematica 8 we have launched the first production example. It is certainly not the end of the story, but I think it's a really good beginning. And I know that even as an expert Mathematica programmer, I've started routinely using natural language input for certain steps in writing programs.

One can also specify programs in natural language to apply to things one has constructed in Mathematica. And in a Mathematica session, one can discard the natural language and just use the generated code by clicking that code. Some interesting examples are shown above.

Now, of course, there are many issues – for example about disambiguation. But the good news is that we've got schemes for addressing these that we've been able to test out well in Wolfram|Alpha.

I have to say that something I thought would be a big issue is the vagueness of natural language. That one particular natural language input might equally well refer to many different precise programs.

And I had imagined it would be a routine thing to have to generate test examples for the user in order to be able to choose between different possible programs.

But in reality this seems to be quite rare: there is usually an "obvious" interpretation, that in typical Wolfram|Alpha style, one can put first, with the less obvious interpretations a click away.

So how well does this all work? We have built out some particular areas of program functionality, and we will progressively be building out many more as time goes on.

They are primarily set up to work in Mathematica. But actually you can see most of them in some form just on the Wolfram|Alpha website – though obviously no references to variables or other parts of a Mathematica session can be used.

How robust is it all? It's definitely usable, but I would certainly like it to be more robust – and we will be working hard in that direction.

One issue that we have faced is a lack of linguistic corpora in the area. We've scoured a couple of decades of our own tech support logs, as well as

Number of bits in a byte. Side length of a chess board. Atomic number of oxygen.

many programming forums, to try to find natural language descriptions matched with precise programs. But we haven't be able to apply anything like the same level of automatic filtering to this process as we've been able to apply in many other areas of "linguistic discovery" for Wolfram|Alpha.

There are zillions of fascinating research projects to do in figuring out generalized grammars for specifying different kinds of programming constructs in natural language – and I will look forward to seeing this field of inquiry develop.

We now have another important source of data: actual examples of natural language programming being done in Mathematica. And looking at our real-time monitoring system for the Wolfram|Alpha server infrastructure, I can see that very soon we are going to have a lot of data to study.

How far will it be possible to get with natural language programming? Even six months ago I thought it was only going to be possible to do fairly simple examples. But seeing what we have actually been able to build, I am extremely optimistic about what will be possible.

The hope would be that in the end one will just have to describe in natural language the goal for one's program – and then an actual program that achieves that goal will be synthesized. Sometimes this will directly be possible from understanding the specification of the goal. Sometimes to create the necessary program will require a whole program-creation process – probably often involving searching for an appropriate program in a space of possible programs, in the style of *A New Kind of Science*.

It will be important to do program simplification – again often achieved by program search – in order to be able to get the simplest and most readable (and perhaps the most efficient) program that meets the requirements that have been given.

At this point, I am still concerned about how much of this will be possible in "interactive times" of a few seconds. But if history is a guide, with good algorithms and heuristics, and a healthy dose of large-scale parallelism, it'll gradually be possible to get the times down.

So what will be the result? I expect natural language programming will eventually become ubiquitous as a way of telling computers what to do. People will be able to get started in doing programming-like tasks without learning anything about official "programming" and programming languages: they'll just converse with their computers as they might converse with another person.

This article is reprinted from his blog at blog.stephenwolfram.com with kind permission of Stephen Wolfram.

What will happen to programming languages? Actually, I think they'll become much more visible and widely known than ever before. Because in natural language programming interfaces one will probably be shown the programming language code that's being synthesized.

People will see that, and gradually learn cases where it's much faster and more precise just to enter code like that directly, without going through natural language.

By the way: in Mathematica we are beginning to have code generation capabilities for low-level languages like C. So it's going to be technically possible to go all the way from natural language input down to something like C. And for some practical purposes – especially with embedded systems – that will no doubt be quite useful.

But when it comes to doing traditional programming alongside natural language programming, there's going to be a great premium on having a succinct readable programming language – like Mathematica.

With the free-form linguistics of Mathematica we are at the first step in a long journey. But it is a journey I'm now confident we can take. After so many years, the science-fiction concept of being able to tell a computer what to do by using plain human language is gradually going to become reality – in a way that fascinatingly coexists with what's been achieved in high-level computer languages.

Fortress and Sun, by Paul Klee

10 Base of our number system. Sum of the first three primes, first four integers and first four factorials.

Pentaplexity

Sir Roger Penrose

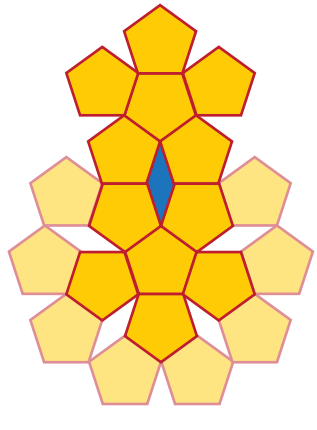
First published in issue 39, 1978

Certain shapes, when matched correctly, can form a tiling of the entire plane but in a non-periodic way. These tilings have a number of remarkable properties, and I shall give here a brief account explaining how these tiles came about and indicating some of their properties.

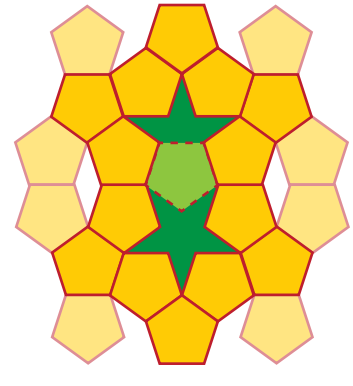
The starting point was the observation that a regular pentagon can be subdivided into six smaller ones, leaving only five slim triangular gaps. This is familiar as part of the usual "net" which folds into a regular dodecahedron, as shown in Figure 1. Imagine now, that this process is repeated a large number of times, where at each stage the pentagons of the figure are subdivided according to the scheme of Figure 1. There will be gaps appearing of varying shapes and we wish to see how best to fill these. At the second stage of subdivision, diamond-shaped gaps appear between the pentagons (Figure 2). At the third, these diamonds grow "spikes", but it is possible to find room, within each such "spiky diamond", for another pentagon, so that the gap separates into a star (pentagram) and a "paper boat" (or Jester's cap?) as shown in Figure 3. At the next stage, the star and the boat also grow spikes, and, likewise, we can find room for new pentagons within them, the remaining gaps being new stars and boats (as before). These subdivisions are shown in Figure 4.

Since no new shapes are now introduced at subsequent stages, we can envisage this subdivision process proceeding indefinitely. At each stage, the scale of the shapes can be expanded outwards so that the new pentagons that arise become the

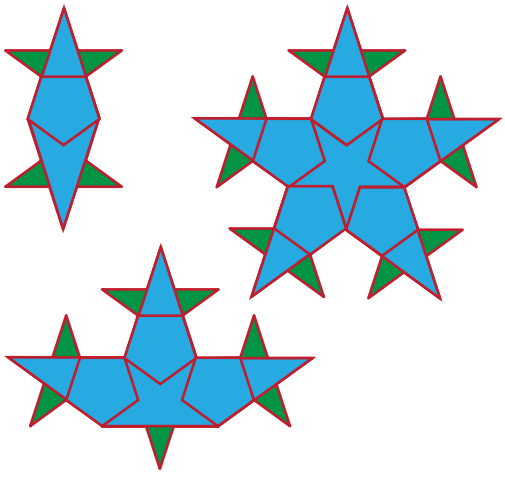
▲ Figure 1



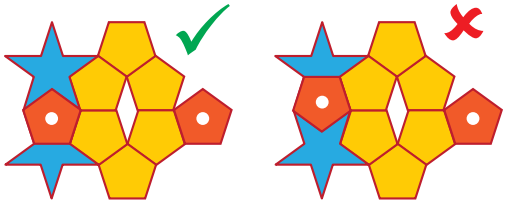
▲ Figure 2



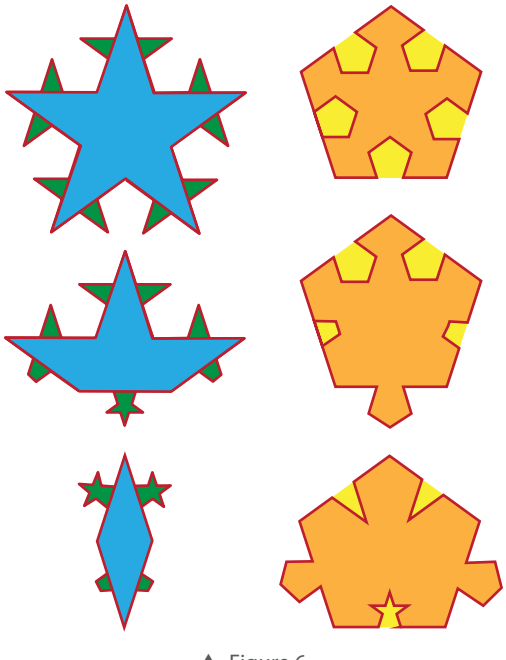
▲ Figure 3



▲ Figure 4



▲ Figure 5



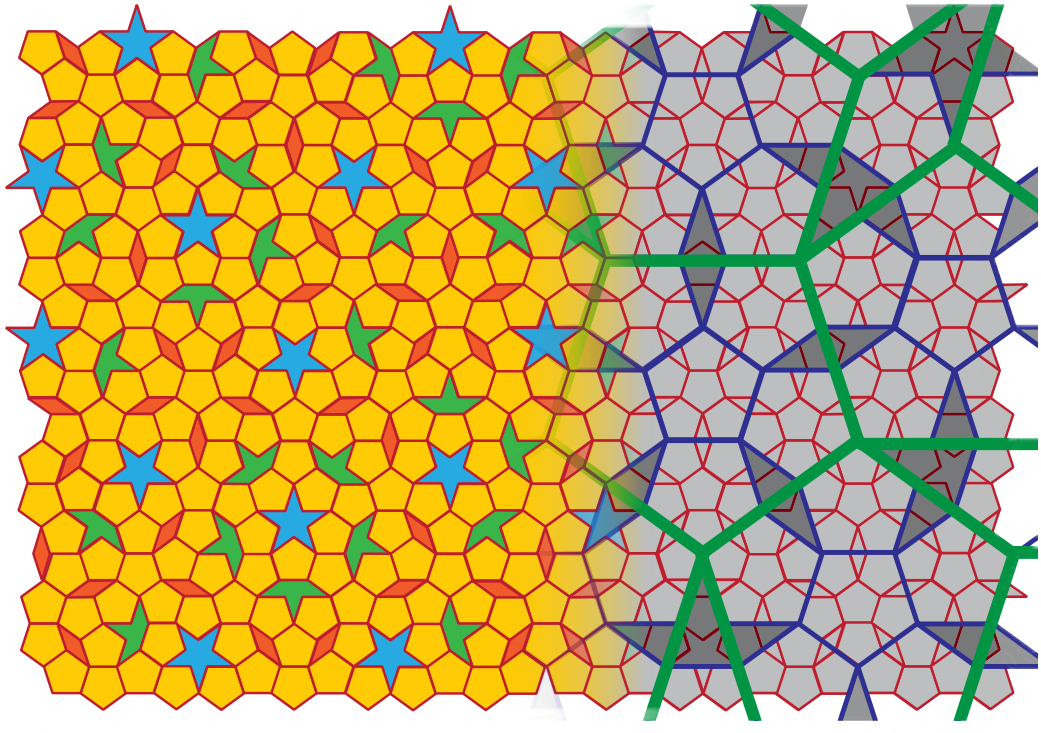
▲ Figure 6

same size as those at the previous stage. As things stand, however, this procedure allows ambiguity that we would like to remove. The subdivisions of a "spiky diamond" can be achieved in two ways, since there are two alternate positions for the pentagon. Let us insist on just *one* of these, the rule being that given in Figure 5. (When we examine the pattern of surrounding pentagons we necessarily find that they are arranged in the type of configuration shown in Figure 5.) It may be mentioned that had the opposite rule been adapted for subdividing a "spiky diamond", then a contradiction would appear at the next stage of subdivision, but this never happens with the rule of Figure 5.

This procedure, when continued to the limit, leads to a tiling of entire plane with pentagons, diamonds, boats and stars. But there are many "incorrect" tilings with the same shapes, being not constructed according to the above prescription. In fact, "correctness" can be *forced* by adopting suitable matching rules. The clearest way to depict these rules is to modify the shapes to make a kind of infinite jigsaw puzzle, where a suggested such modification is given in Figure 6. It is not hard to show that any tiling with these six shapes is forced to have a hierarchical structure of the type just described.

Properties of these Tilings

Furthermore, the forced hierarchical nature of this pattern has a number of very remarkable properties. In the first place, it is necessarily non-periodic (i.e. without any period parallelogram). More about this later. Secondly, though the completed pattern is not uniquely determined – for there are $2^{\aleph_0}$ different arrangements – these different arrangements are, in a certain "finite" sense, all indistinguishable from one another! Thus, no matter how large a finite portion is selected in one such pattern, this finite portion will appear somewhere in *every* other completed pattern (infinitely many times, in fact). Thirdly, there are many unexpected and aesthetically pleasing features that these patterns exhibit (see Figure 7). For example, there are many regular decagons appearing, which tend to overlap in places. Each decagon is surrounded by a ring of twelve pentagons, and there are larger rings of various kinds also. Note that every straight line segment of the pattern extends outwards to infinity, to contain an infinite number of line segments of the figure. The hier-



▲ Figure 7

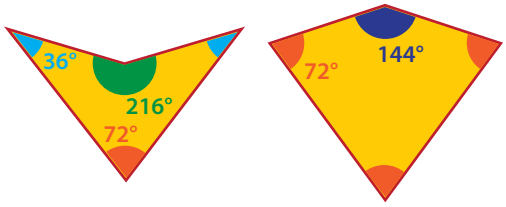
▲ Figure 8

archical arrangement of Figure 7 is brought out in Figure 8.

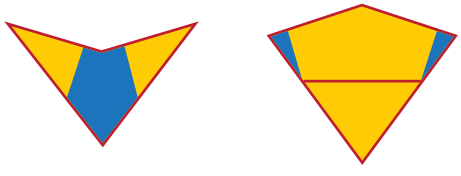
After I had found this set of six tiles that forces non-periodicity, it was pointed out to me (by Simon Kochen) that Raphael Robinson had, a number of years earlier, also found a (quite different) set of six tiles that forces non-periodicity. But it occurred to me that with my tiles one can do better. If, for example, the third "pentagon" shape is eliminated by being joined at two places to the "diamond" and at one place to the bottom of the "boat", then a set of *five* tiles is obtained that forces non-periodicity. It was not hard to reduce this number still further to four. And then, with a little slicing and rejoining, to *two*!

The two tiles so obtained are called "kites" and "darts", names suggested by John Conway. The precise shapes are illustrated in Figure 9. The matching rules are also shown, where vertices of the same colour must be placed against one another. There are many alternative ways to colour these tiles to force the correct arrangement. One way brings out the relation to the pentagon-diamond-boat-star tilings shown in Figure 10. A patch of assembled tiles (partly coloured in this way) is shown in Figure 11. The hierarchical nature of the kite-dart tilings can be seen directly, and is illustrated in Figure 12. Take any such tiling and bisect each dart symmetrically with a straight line segment. The resulting half-darts and kites can then be collected together to make darts and kites on a slightly larger scale: two half-darts and one kite make a large dart; two half-darts and two kites make a large kite. It is not hard to convince oneself that every correctly matched kite-dart tiling is assembled in this way. This "inflation" property also serves to prove non-periodicity. For suppose there were a period parallelogram. The corresponding inflated kites and darts would also have to have the same period parallelogram. Repeat the inflation process many times, until the size of the resulting inflated kites and darts is greater than that of the supposed period parallelogram. This gives a contradiction.

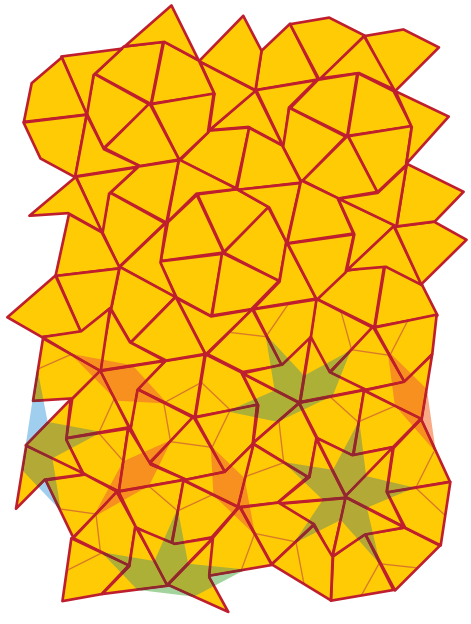
The contradiction with periodicity shows up in another striking way. Consider a very large area containing d darts and k kites, which is obtained referring to the inflation process a large number of times. The larger the area, the closer the ratio $x = k/d$ of kites to darts will be to satisfying the recurrence relation $x = (1 + 2x)/(1 + x)$, since, on inflation, a dart and two kites make a larger kite, while a dart and a kite make larger dart. This gives, in the limit of an infinitely large pattern,



▲ Figure 9



▲ Figure 10



▲ Figure 11

$x = \frac{1}{2}\left(1+\sqrt{5}\right) = \varphi$, the golden ratio! Thus we get an *irrational* relative density of kites to darts – which is impossible for a periodic tiling. (This is the *numerical* density. The kite has φ times the area of the dart, so the total area covered by kites is φ^2 $(= 1 + \varphi)$ times that covered by darts.)

Jigsaws and beyond

There is another pair of quadrilaterals which, with suitable matching rules, tiles the plane only non-periodically: a pair of rhombuses as shown in Figure 13. A suitable shading is suggested in Figure 14, where similarly shaded edges are to be matched against each other. In Figure 15, the hierarchical relation to the kites and darts is illustrated. The rhombuses appear mid-way between one kite-dart level and the next inflated kite-dart level.

Many different jigsaw puzzle versions of the kite-dart pair or the rhombus pair can evidently be given. One suggestion for modified kites and darts, in the shape of two birds, is illustrated in Figure 16.

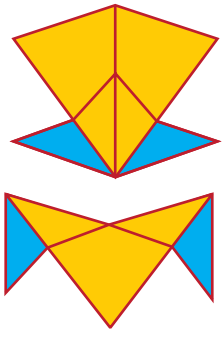
Other modifications are also possible, such as alternative matching rules, suggested by Robert Ammann (see Figure 17) which force half the tiles to be inverted.

Many intriguing features of these tilings have not been mentioned here, such as the pentagonally-symmetric rings that the stripes of Figure 14 produce, Conway's classification of "holes" in kite-dart patterns (i.e. regions surrounded by "legal" tilings but which cannot themselves be legally filled), Ammann's three-dimensional version of the rhombuses (four solids that apparently fill space only non-periodically), Ammann's and Conway's analysis of "empires" (the infinite system of partly disconnected tiles whose positions are forced by a given set of tiles). It is not known whether there is a *single* shape that can tile the Euclidean plane non-periodically. For the hyperbolic (Lobachevski) plane a single shape can be provided which, in a certain sense, tiles only non-periodically (see Figure 18) – but in another sense a periodicity (in one direction only) can occur. (This remark is partly based on suggestions of John Moussouris.)

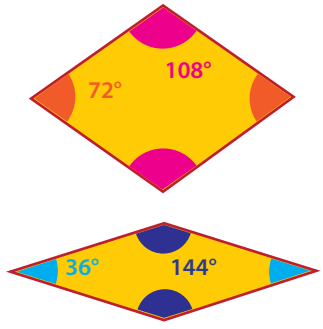
References

1. M. Gardner, Scientific American, January 1977, pp. 110–121
2. R. Penrose, Bull. Inst. Maths. & its Applns. 10, No. 7/8, pp. 266–271 (1974)

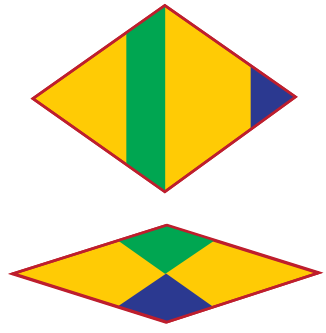
Open meandric number. These are the number of meanders in non-self-intersecting oriented curves.



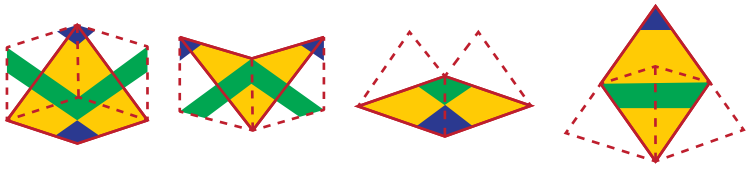
▲ Figure 12



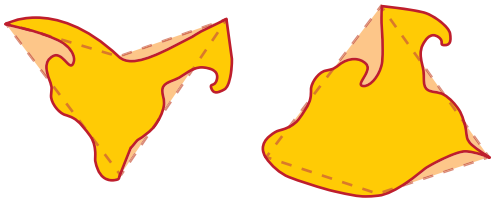
▲ Figure 13



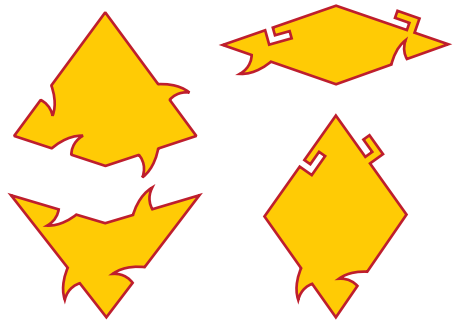
▲ Figure 14



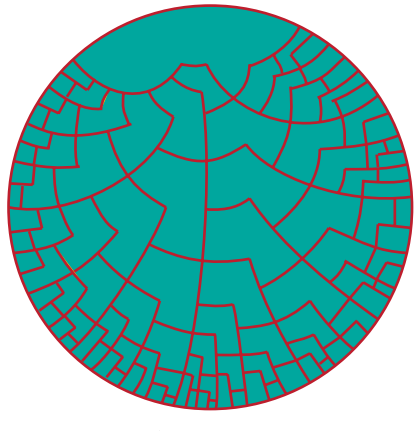
▲ Figure 15



▲ Figure 16



▲ Figure 17



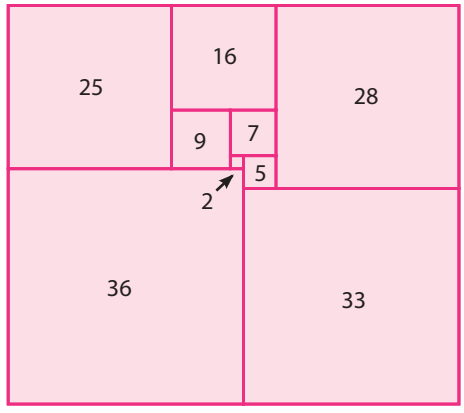
▲ Figure 18

Squared Squares

Philipp Kleppmann, Corpus Christi

Figure 1 shows a rectangle that is dissected into smaller squares, all of which have different sidelengths. Such rectangles are called *squared rectangles*. Of course, rectangles like this one can be constructed by trial and error if you have enough time or a computer. The task becomes harder if you try to produce a squared square. The challenge of finding one arose in the early twentieth century from a problem in a mathematical puzzle book called The Canterbury Puzzles [5]. It wasn't even clear that a squared square existed, until R. Sprague found one in 1939 [3], more than 30 years later.

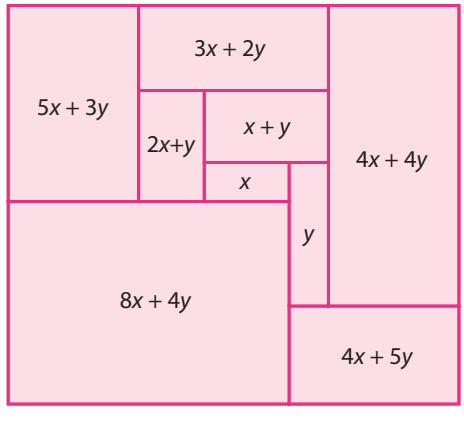
In the 1930s, the four Cambridge undergraduates Roland Brooks, Cedric Smith, Arthur Stone, and William Tutte came across this problem and devised some very clever methods of producing squared rectangles and squares using the theory of electrical networks, some of which I will present here. The present-day logo of the Trinity Mathematical Society is a squared square, in recognition of the four Trinity students.



▲ **Figure 1** A squared rectangle

The low-tech method

Draw a rectangle cut up into smaller rectangles, as in Figure 2. Squint at it and imagine that it is just a bad drawing of a squared rectangle. Assign values x and y to the sidelengths of two of the `squares' as shown in the figure. From these it is easy to determine all other sidelengths: First $x + y$ above the two starting squares, then $2x + y$ to the left, and so on. We have to make sure that the two vertical sides of the big rectangle have the same length. For this we need $(5x + 3y) + (8x + 4y) = (4x + 4y) + (4x + 5y)$, i.e. $5x = 2y$. So, taking $x = 2$ and $y = 5$, we get the squared rectangle in Figure 1.



▲ **Figure 2** A badly drawn squared rectangle

This method was used by Arthur Stone to construct his first squared rectangle [4]. While it is easy to apply, it is also luck-dependent. You can't count on finding a dissection that can become a squaring, and sometimes the equations give negative values for some sidelengths. This makes a systematic analysis very difficult. For example, if one is interested in the smallest number of squares that a rectangle can be cut up into, it is not at all clear how to show that there are no smaller ones. In fact, the smallest number is 9. This was proved in [1] using the following more refined method.

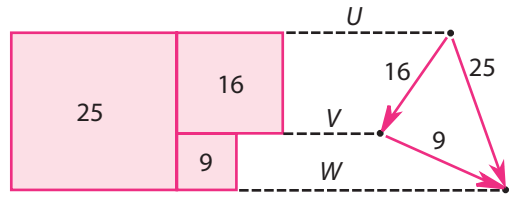
The high-tech method

Suppose we have a squared rectangle, such as the one in Figure 1. We construct a directed graph with a vertex for each horizontal line segment and an edge for each square. There is an arrow from a vertex v to a vertex w if and only if the corresponding horizontal line segments V and W in the rectangle are connected by a square and V is above W. We label the arrow with the sidelength of this square. Figure 3 illustrates the procedure.

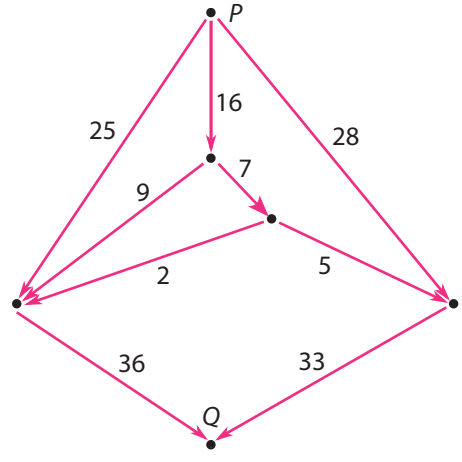
The graph in Figure 4 is constructed in this way from the rectangle in Figure 1. It is called the Smith diagram of the rectangle.

Now call P and Q the poles of the network, and interpret the labels of the edges as currents. There are a couple of things in the graph that you may notice:

1. For any vertex that isn't a pole, the sum of currents entering it is equal to the sum of currents flowing out of it: The sum of the sidelengths of squares lying directly above one horizontal line segment in the squared rectangle is the same as the sum of the sidelengths of squares lying directly below it.
2. The sum of currents around any circuit is zero (counting currents in the 'wrong' direction as negative currents). This is because the length of any straight vertical path from one horizontal line segment to another one is the same, no matter which squares it passes through.
3. The sum of the currents leaving P is equal to the sum of the currents entering Q, since the lengths of the two horizontal sides of the rectangle are equal.



▲ **Figure 3** How to make a Smith diagram



▲ **Figure 4** The full Smith diagram

And – hey presto! – we've built an electric network in which the given currents are valid as long as we assume that each wire has unit resistance. (1) and (2) are called Kirchhoff's laws.

In fact, this construction works in the other direction as well: If we construct an electric network satisfying the three conditions above and which has different currents along all of its wires, then it is a blueprint for a squared rectangle! Of course, the network encapsulates the same information as the squared rectangle, but it has many advantages over the first method. Graphs are well-established mathematical objects, so we can fall back on a large body of theory. In particular, the theory of electric networks can be used for further investigations. See [1] for more.

Graphs can be searched systematically. For example, to show that there are no squared rectangles consisting of fewer than 9 squares we can search all directed graphs with at most 8 (and at least 2) edges and try to assign distinct values to the edges in such a way that the three conditions are satisfied. But there aren't any [1]! The rectangle in Figure 1 consists of 9 squares, so it is proved that the smallest number of squares in a squared rectangle is 9.

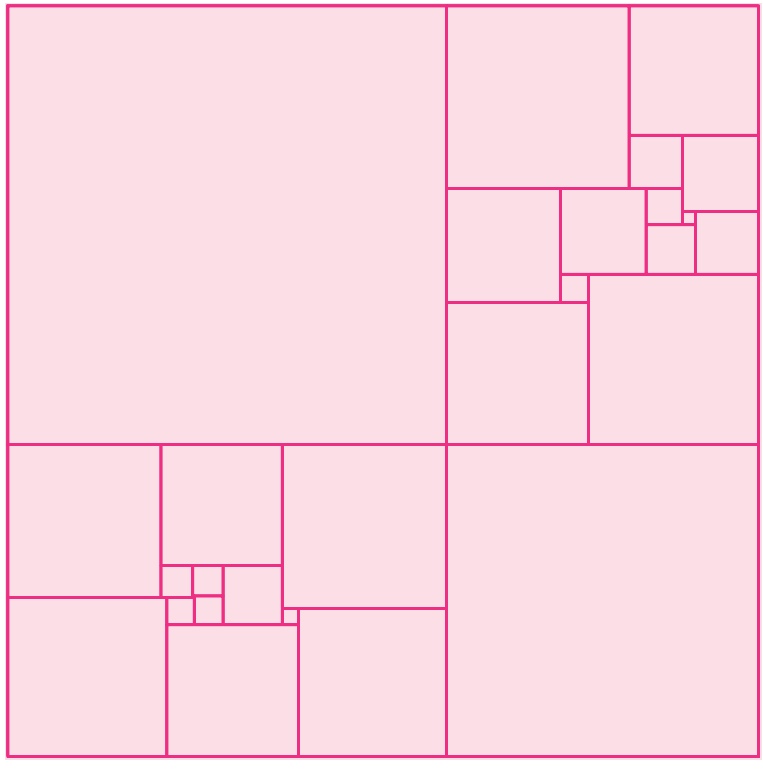
A squared square

One way of finding a squared square is to exhaustively search through all squared rectangles, until you spot one. The smallest has 21 pieces [3], so this might take a while. It turns out that looking for a squared rectangle that can be cut up into squares in two completely different ways (meaning that none of the squares of one dissection appear in the other dissection) will get you there a lot faster.

The smallest such rectangle is 422 × 593 and can be cut up into 13 pieces in two different ways [4]. The sidelengths of the component squares are 18, 38, 49, 67, 72, 85, 103, 116, 154, 175, 192, 230, 247 and 2, 22, 37, 39, 41, 43, 80, 164, 178, 200, 207, 215, 222, respectively. The two rectangles are combined with two squares to form one large squared square, as shown in below. You might like to assign the sidelengths to the component squares yourself!

References

1. R.L. Brooks, C.A.B. Smith, A.H. Stone, W.T. Tutte, *The Dissection of Rectangles into Squares*, Duke Math. J. 7, pp. 312-340 (1940)
2. A.J.W. Duijvestijn, *Simple perfect squared square of lowest order*, J. Combin. Theory Ser. B 25, pp. 240-243 (1978)
3. R. Sprague, *Beispiel einer Zerlegung des Quadrats in lauter verschiedene Quadrate*, Math. Zeitschrift 45, pp. 607-608 (1939)
4. W.T. Tutte, *Squaring the Square*, http://www.squaring.net/ history_theory/brooks_smith_stone_tutte_II.html
5. http://www.squaring.net/sq/ss/ss.html




This is how we order lunch:
$L^* = \arg\max_{L \in \mathcal{L}} U_{\text{CANTAB}}(X_0(L), \ldots, X_{N-1}(L))$
interested in being the Nth?
cantabcapital.com/yourfuture
CANTAB
CAPITAL PARTNERS

Tricky Teacups

Vito Videtta, Trinity Hall

Here is a puzzle I found when volunteering at maths outreach events in Cambridge: it is called *Aunty's Teacups*. We are given 16 teacups, four each in each of four colours, and, similarly, 16 saucers. Arrange the cups on top of the saucers in a 4 by 4 square grid so that:

1. in each row and column, there is one cup of each colour;
2. in each row and column, there is one saucer of each colour;
3. (Orthogonality Condition) no cup-saucer colour combination is repeated. (To be clear, this means that, for example, red-on-green and green-on-red are both allowed.)

I was rather intrigued when I first saw this puzzle; it appeared to be so simple, yet everybody who tried it quickly found out it was a Pandora's Box. It was apparent that I was hooked as soon as I got home; I immediately started working on a solution. Before we get to that though (and to give you a chance to try it for yourself), let's go through some of the history of this problem.

Square arrangements of the above type were first studied by Leonhard Euler. In a seminal paper published in 1782, he poses the following problem: 'Given a group of 36 officers of six different ranks, one each from six different regiments, is it possible to arrange the officers in a square, in such a manner that in each line, vertical or horizontal, there is one officer of each rank and one from each regiment?' This problem came to be known as "Le Problème des 36 Officiers"; despite its apparently simple formulation, it turned out to be one of the hardest mathematical problems ever posed. Consequently, this led to a barrage of new mathematics being created as more and more mathematicians tried to rise up to Euler's Challenge.

As with much great mathematics, this particular problem managed to find its way into popular culture in the form of many puzzles. One version that is particularly simple to set up requires a pack of cards. Take the Ace, King, Queen and Jack of each suit. Arrange the cards in a 4 by 4 square grid so that in each row and column, there is one card of each rank and one of each suit. You can easily see that this puzzle is equivalent to the Teacups problem (in fact, it is slightly easier, since we don't have to worry about the orthogonality condition).

Journey towards a solution

A *Latin square of order n* is an $n \times n$ matrix containing n copies of the numbers 1 to n arranged so that in each row and column, each number appears once and only once. Latin squares arise naturally as the multiplication tables of finite groups (of course, the symbols we use to label the square are unimportant). Suppose A and B are Latin squares of order n; we say that A and B are *orthogonal* if (A_{ij}, B_{ij}) ranges through all possible ordered pairs as i, j range through all legal indices. Moreover, we call B an orthogonal mate to A. In a sense, orthogonal Latin squares (OLSs) are Latin squares that are as different as possible from each other.

We now see how to translate the Aunty's Teacups problem into mathematics; we are merely searching for a pair of OLSs of order 4.

The number of quarter or half turns required to optimally solve a Rubik's cube in the worst case.

I began my investigation by, rather shamefully, writing a computer program to find all possible solutions. I started by labelling the four colours using the numbers 1 to 4 and using the notation (c, s) to describe the entries in the 4-by-4 matrix, where c is the cup colour and s is the saucer colour. After five hours of coding and a split second of computing, it spat out a completely unintelligible sequence of the numbers 1 to 4 and, somewhat more importantly, the number of solutions it had found: 6912. That was nice, but uninformative.

And so I left the problem there for a while, in which time I finished my first two years at University. But when I showed the puzzle to a young girl and her father at another outreach event, an obvious fact hit me with more force than a meteor strike: given any solution, I could cyclically permute the rows and columns to generate new solutions. Topologically then, the grid can be wrapped around into a tube and the ends connected to make a doughnut; a sort of "Teacup Torus".

In fact, I could act on the solution space X by the group $S_4 \times S_4$ via row and columns swaps on any solution. I called this group the Automorphism group of X. (This terminology is non-standard; I chose it purely by analogy with Galois theory.) This allowed me to define an equivalence relation $\sim$ on the space of solutions with the equivalence classes precisely the orbits of the action: we say $x \sim y$ iff there is $\sigma \in S_4 \times S_4$ such that $x = \sigma(y)$ where $\sigma = (\sigma_1, \sigma_2)$ acts on y by permuting the rows via σ_1 and the columns via σ_2. Having defined this equivalence relation, it then became very natural to ask if there was a convenient choice of representative for each equivalence class. The answer here is 'yes'; by permuting rows and columns, I could always transform any given solution to the form

(1, 1)	(2, ?)	(3, ?)	(4, ?)
(2, ?)	?	?	?
(3, ?)	?	?	?
(4, ?)	?	?	?

Moreover, the action has the following highly desirable property: if $x, y \in X$ are distinct solutions in the same orbit, then there is a unique $\sigma \in S_4 \times S_4$ such that $x = \sigma(y)$. Hence, each equivalence class has size $(4!)^2 = 576$ and so this gives $6912/576 = 12$ different equivalence classes. This was a major step towards the solution, but I was still not satisfied; 12 was "too big" and I felt that I could quotient out more group actions from the solution space.

Again, progress fell silent as I was revising for my final exams. The final piece of the puzzle came to me as I walked back to college along Burrell's Walk one evening early in May. I had been revising Galois theory, which got me thinking about groups and (inevitably) a certain puzzle involving pieces of fine china. I was thinking about how I could incorporate permutations of colour into my solution; clearly, I had completely exhausted all possible permutations of rows and columns and so colour permutations stood as the final frontier. The problem I faced was that in the top-left hand corner of the above scheme, colour 1 had already been fixed in the second coordinate. It occurred to me while I was walking that, actually, this did not present a problem: *just permute the remaining*

colours in the second coordinate among themselves. By doing so, I could force any solution to assume the form

(1, 1)	(2, ?)	(3, ?)	(4, ?)
(2, 2)	?	?	?
(3, 3)	?	?	?
(4, 4)	?	?	?

Hence, I had the final automorphism group: $\mathrm{Aut}(X) = S_4 \times S_4 \times S_3$, a group of size 3456. It remained to find representatives for the equivalence classes. After a bit of searching, they presented themselves:

$x_1 =$

(1, 1)	(2, 3)	(3, 4)	(4, 2)
(2, 2)	(1, 4)	(4, 3)	(3, 1)
(3, 3)	(4, 1)	(1, 2)	(2, 4)
(4, 4)	(3, 2)	(2, 1)	(1, 3)

$x_2 =$

(1, 1)	(2, 4)	(3, 2)	(4, 3)
(2, 2)	(1, 3)	(4, 1)	(3, 4)
(3, 3)	(4, 2)	(1, 4)	(2, 1)
(4, 4)	(3, 1)	(2, 3)	(1, 2)

This was exactly what I was expecting: there are two equivalence classes resulting from this action. Also, the above solutions lie in different orbits, which preserves the uniqueness of action. Hence, I had $2 \times 3456 = 6912$ solutions altogether, which agreed with my computer search.

Designing Experiments

All of the above discussion has a distinctly pure flavour and may have made some of my applied readers slightly nauseous. Fear not my friends, we now present an application of the above theory.

Suppose that the brilliant genius Prof. Tarquin Walter Kornman is organising examinations for his students at Camford University. He is currently arranging an exam timetable for his four students, Alice, Daniel, Grace and Timothy, who must each take four papers. Tarquin aims to design a timetable that is as efficient as possible, but considers it a form of cruelty to force a student to sit more than one paper per day. However, in spite of his compassion, his genius lends him certain eccentricities, chief among which is an extreme desire to eliminate bias as far as possible. He does so by choosing, for each day, four different examination start times. Given this information, how should Tarquin finish the exam timetable?

Here, Tarquin has four collections of symbols, or *treatments*, to deal with: Start Time, Exam Date, Student Name and Paper Number. It turns out that the most efficient and unbiased arrangement uses a pair of OLSs of order 4, preferably chosen at random. All that Tarquin need do is to specify which treatment will label the rows, which will label the columns and finally to use the other two treatments as sets of symbols with which to fill the square. For example, one solution might be to use the Student Names as row labels, the Exam Dates as column labels and then to fill the table with the Paper Numbers and Start Times. A similar strategy will work for any similar experiment; such arrangements are called *Pairwise Balanced Designs* and are frequently used to ensure efficiency and elimination of bias when designing an experiment.

Now, suppose that Tarquin is dissatisfied with the way the examinations were run this year. He believes that the students had not been examined thoroughly enough, so decides to make the exams harder by introducing two new papers. By sheer luck, two new students join the course the following year, so that now we have six papers to be taken by six students over six days, each at six different starting times. Tarquin, now knowing that the key lies in orthogonal latin squares, proceeds to construct a solution. However, after many long and painful attempts, he is reduced to a blubbering wreck on his office floor, since he is unable to find one. He has convinced himself of his own stupidity, but he need not be so harsh on himself…

Euler's Conjecture: or, Tarquin redeemed

Earlier I mentioned the famous "Problème des 36 Officiers" that was first stated by Euler. Like Tarquin, Euler devoted much effort to solving this problem, but was also unable to find a solution. Euler had already developed techniques to construct pairs of Orthogonal Latin Squares of order n for n odd or divisible by 4. Given the (trivial) impossibility of the case $n = 2$ and his unsuccessful attempts for $n = 6$, Euler made a bold claim: There do not exist a pair of Orthogonal Latin Squares of order $4t + 2$, for any $t \geq 0$. One can only imagine Euler's reasoning for making such an extreme claim on the back of two pieces of evidence; I daresay it smacks of Physicists' Induction.

A first cry of success came in 1900, when Gaston Tarry published a proof that the case $n = 6$ is in-

deed impossible. Tarry was born in France and moved to Algeria to work as an administrator. Although an amateur, he had an amazing capacity for combinatorial problems. His proof proceeds as follows: he began by reducing all order 6 Latin Squares to 17 types, via careful and painfully detailed reasoning about cycle types in S_6. From there, he reasoned that for a solution to exist, we must find such a Latin Square that possesses a complete set of transversals. A transversal is a subset of n cells within an $n \times n$ square, labelled with the symbols 1 to n, such that in each row and column of the square there is a cell of the subset. At that point, he took a minor detour to give an interesting study of the $n = 4$ case using transversals. Returning to the proof, he managed to reduce to 3 the number of Latin Squares that need to be considered by showing that some cells of some Latin Squares cannot be part of any transversal. Finally, he considered these three cases in turn, using a greedy algorithm which he calls The Method of Order to show that no orthogonal mate can exist for any of them. While we can be glad that this proof resolves the problem, it does lack a certain panache; it tells us that no orthogonal mate can exist, but doesn't tell us *why* it can't exist. Indeed, even Tarry was disappointed by this; he writes, "The method of order, which does not shed any light on the problems it resolves, should not be used unless we cannot do otherwise; it is a last resort".

Since Tarry's proof was published, shorter and more informative proofs have been found. See, for example, Fisher and Yates (1934), Yamamoto (1954) or Stinson (1984). Stinson's proof rather interestingly highlights a link between Latin Squares and coding theory!

The Fall of Euler's Conjecture

Following the proof for the case $n = 6$, the subject seemed to have died down slightly. This changed in May 1959, when Raj Bose and Sharadchandra Shrikhande managed to construct a pair of Orthogonal Latin Squares of order 22, thus disproving Euler's Conjecture. In fact, it didn't stop there; shortly after that paper was published, Ernest Parker published a paper in which he presented an example with $n = 10$. Euler's Conjecture was crumbling fast; in fact, he couldn't have been more wrong. In one final paper published by Bose and Shrikhande in 1960, they proved that there exist a pair of OLSs of order $4t + 2$ for infinitely many t. To add insult to injury, the proof didn't even use very advanced techniques; it relied on fairly straightforward properties of finite fields and techniques from Combinatorial Design theory.

The sudden success with which Euler's Conjecture was disproved sparked a new wave of interest in Latin Squares. To capitalise on this interest, Józef Dénes and Donald Keedwell wrote a comprehensive volume on Latin Squares in 1974. It became an instant hit; such was the interest in Combinatorial Designs that a sequel was published 17 years later. Between them, the two volumes cover many aspects on the theory and applications of Latin Squares and contain no less than 4 complete chapters devoted to the idea of Orthogonality. Applications included Experimental design, Statistics, Error-correcting codes, Algebra and Geometry, to name only a few.

References, Further Reading

My sincerest thanks go to Sue Hickman Pinder from the Millennium Mathematics Project and to Rebecca Paul for proofreading.

1. I. Anderson, *Combinatorial Designs and Tournaments*, Oxford University Press (1997)
2. R. C. Bose and S. S. Shrikhande, *On the falsity of Euler's conjecture about the non-existence of two orthogonal latin squares of order 4t + 2*, Proceedings of the National Academy of Sciences (1959)
3. R. C. Bose and S. S. Shrikhande, *On the construction of sets of mutually orthogonal latin squares and the falsity of a conjecture of Euler*, Trans. of the American Math. Society (1960)
4. J. Dénes and A. D. Keedwell, *Latin Squares and their Applications*, English Uni. Press (1974)
5. J. Dénes and A. D. Keedwell, *Latin Squares: New Developments in the Theory and Applications*, Elsevier Science Publications (1991)
6. M. Gardner, *Sphere Packing, Lewis Carroll and Reversi*, CUP (2009)
7. D. R. Stinson, *A Short Proof of the Nonexistence of a Pair of Orthogonal Latin Squares of Order Six*, Journal of Combinatorial Theory (1984)
8. G. Tarry, *Le Problème des 36 officiers*, Comptes Rendu de l'Association Française pour l'Avancement de Science Naturel (1901)
9. *nrich.maths.org/32/index*
10. *www-history.mcs.st-and.ac.uk/Biographies/Tarry.html*

David Phillips, Alec Barnes-Graham

Archimedeans Annual Problems Drive

1 Dazzling Dice

You roll a (standard fair six-sided with sides 1 to 6) die an infinite number of times, recording the total score attained so far after each roll as a sequence. What are the most and least likely numbers to appear in this sequence, and with what probabilities do they occur?

2 Snappy Surds

Find all integers such that

$$\sqrt{n - 4\sqrt{n - 19}}$$

is also an integer.

3 Painful Primes

Exactly one of the following numbers is prime. Which one?

852,081	1,050,589
967,535	1,052,651
999,917	1,073,254
999,919	1,093,411

4 Compelling Convergence

Determine whether the following series converge or diverge, and determine the value of any that converge.

$$\sum_{n=1}^{\infty} \left| \frac{\sin n}{n^2} \right|^{1/2} \qquad \sum_{n=1}^{\infty} \frac{2 \sin n}{n^2} \left| \frac{2(1 - \sin^2 n)}{1 - \cos 4n} \right|^{1/2} \qquad \sum_{n=1}^{\infty} \frac{(\log 2)^{2n+1}}{2n + 1}$$

Only non-trivial integer n with the property that $1^2 + \ldots + n^2$ is a perfect square, in this case 70^2.

5 Superb Sets

For each of the following sets, determine whether it is finite, countable or uncountable. Give the explicit sizes of the finite sets, and for any uncountable set, determine whether it bijects with $\mathbb{R}$.

- Group homomorphisms $(\mathbb{Z},+) \to (\mathbb{Q},+)$
- Group homomorphisms $(\mathbb{Q},+) \to (\mathbb{Z},+)$
- Equivalence relations on $\mathbb{Q}$
- Sequences in $\mathbb{Q}$ that converge to some member of $\mathbb{Q}$

6 Triumphant Treasures

The planet Zog has radius 1 has an associated geostationary moon of negligible radius. You have followed the evil space-pirate Blackmoustache to this system, in which he has buried his treasure. You know that:

i. The moon lies a distance λ from the planet, with $1 < \lambda < 1.5$;

ii. The centre of Zog is denoted Z;

iii. The city of Luna lies on the closest point of the planet to the moon;

iv. The city of Antiluna is antipodal (at the other end of the diameter) to Luna;

v. The treasure lies at least λ away from Antiluna;

vi. The city of Luna produces so much toxic waste that any point P in the planet with $\measuredangle PZL < \alpha$ cannot contain the treasure;

vii. The core of Zog is molten, so the treasure does not lie within it;

viii. If the core has volume V and surface area A, then
$$\tfrac{A}{4\pi} + 6\tfrac{V}{A}\cos\alpha > \lambda^2 - 1;$$

ix. If the treasure lies at the point T, then the following inequality holds: $|TZ|\sin(\measuredangle TZL) < \sqrt{(\lambda^2-1)\sin^2\alpha + \tfrac{1}{4}\sin^2(2\alpha) - \tfrac{1}{2}\sin(2\alpha)}$;

x. The city of Midi lies exactly halfway between Luna and Antiluna, with antipodal city Centra. Then the treasure is known to be in the plane containing Luna, Midi and Antiluna, and to be at least as close to Midi as to Centra.

Where is the treasure buried?

7 Curious Coins

Two players play a game on an $n \times n$ square table on which coins of diameter 1 are placed in turn. The winner is the one who plays the last coin. For which n do you want to play first?

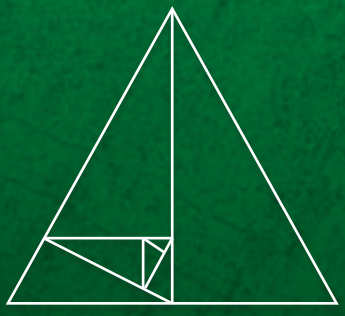
NB: The coins must have their centre above the table, must be placed flat and cannot be stacked.

8 Perceptive Polygons

Begin with an equilateral triangle of side length 1, and draw its circumcircle. About this, circumscribe a square, and then draw the circle around this. Repeat this infinitely many times, each time circumscribing a regular n-gon around the outermost circle, and then drawing the circumcircle of that, forming the new outermost circle. Does this object fit inside a circle of radius 100?

9 Terrible Triangles

In an equilateral triangle with side length 1, consider dropping a perpendicular from a vertex onto the opposite side. Then, repeat this process, spiralling in clockwise, as in the picture. Where (in Cartesian coordinates, calling the bottom left vertex the origin) is the point to which this process converges?



10 Rough Relations

Let R be a relation that is "anti-transitive", that is if aRb and bRc, then cRa. Then, define $f(n)$, for $n \in \mathbb{N}$, as the least $m \in \mathbb{N}$ such that there exists a set T, with $|T| = n$ and $a, b \in T \Rightarrow aRb$ or bRa, so that exactly m unordered pairs $(s, t) \in T \times T$ have the properties:

i. $s \neq t$;

ii. sRt and tRs.

Find, for all $n \in \mathbb{N}$, the value of $f(n)$.

11 Gorgeous Geometry

Let C be the mid-point of OD, and let Q lie on the semicircle through D with centre C, whose diameter is perpendicular to OD. Points A and B lie in the plane of the semicircle, are equidistant from O and also from Q. The point R completes the rhombus $QARB$.

Find the locus of R as Q traverses the semicircle, with the distances OA, OB, QB, AR and BR remaining fixed.

12 Mysterious Matchings

This problem can be solved easily by 5 to 10 year olds:

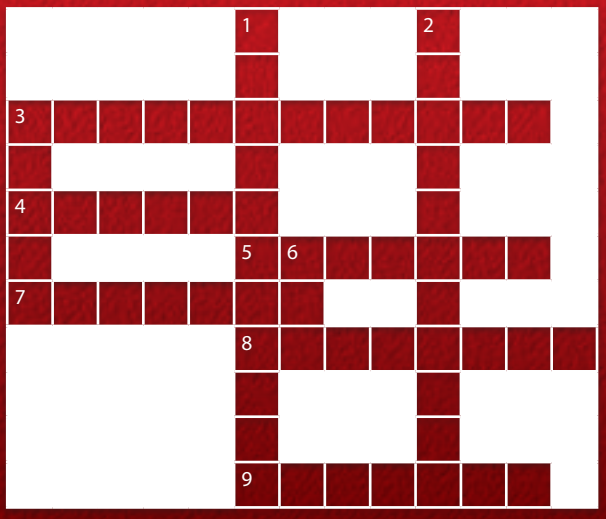
1235 → 0	8738 → 4	0000 → 4	8317 → 2	
1101 → 1	3275 → 0	9834 → 4	2814 → 3	
2222 → 0	2176 → 1	9393 → 2	5656 → 2	3821 → ?
3535 → 0	8261 → 3	7272 → 0	0909 → 4	
9232 → 1	7068 → 4	1818 → 4	7777 → 0	

13 Dazzling Digits

Hardy and Ramanujan are playing a game, where on each turn, Hardy names some digit (which need not be distinct from previous digits), and then Ramanujan inserts it into the expression ∗∗∗∗ − ∗∗∗∗, in place of one of the stars.

Hardy is aiming to maximise the value of the expression, Ramanujan to minimise it. Being Cambridge mathematicians, they both play perfectly. What is the value of the expression?

14 Cryptic Crossword



Across

3 Commuter (BA) pouring ale mixture. (7,5)
4 Not single or nothing! (6)
5 With an angle that small, there's no degrees here. (7)
7 Rain men confused, but still discovering differential geometry. (7)
8 One bash in frontless city – it fixes everything. (8)
9 Three different sides of disc – ale needed. (7)

Down

1 Equate Rn ions containing Hamilton's group. (11)
2 Mix paint on hide of Archimedes' cows. (11)
3 That's sum royal snake! (5)
6 Also intersection. (3)

Solutions to the Archimedeans Problems Drive can be found on page 94.

Pi in Fours

John Conway and Michael Guy

First published in issue 25, 1962

The famous "four 4s problem" asks you to arrange four 4's, and any number of the ordinary mathematical symbols, to give as good an approximation to Pi as you can find.

We shall allow the symbols (,), +, –, × and ÷, the usual notations for roots $\sqrt{}$ and $\sqrt[4]{}$, powers, factorials and the decimal notation 44, .4 and $.\dot{4}$. Pi itself, logarithms and trigonometric functions may not be used. Factorials are to be of integers only, otherwise $\pi = \sqrt{(-\sqrt{4}/4)!^4}$. We shall also not allow such monstrosities as $.\sqrt{4}$.

For example,

$$\sqrt{\sqrt{\left(\frac{4!!+4}{4!!}\right)^{4!!}}}$$

is a very good approximation to e, and can clearly be modified to be as good as we please. It can furthermore be improved so as to only use three 4's, since, as $n \to \infty$, $n / \sqrt[n]{n!} \to e$.

We may derive similar "explicit" formulae for various interesting numbers. Thus $n\sqrt[n]{a} - n \to \log a$, so that we obtain a sequence of approximations to log 2, log 5, and $\log_a b$ for a variety of rational a and b (e.g. $\log_{10}2$ or $\log_{10}3$). Our best result of this kind for π has seven 4's, and is derived from

$$\pi = \lim_{n\to\infty}\left(\frac{2^n\, n!}{\sqrt{\sqrt{n}}\,\sqrt{(2n)!}}\right)^4.$$

We can also find $\log \pi$ in seven 4's, but as yet we have not been able to find any formula of this kind for Euler's constant γ.

We shall now show that the above devices are unnecessary. In fact:

Theorem 1 Any real number may be approximated arbitrarily closely using only four 4's and the usual symbols.

Proof: It follows from the formula $n\left(\sqrt[n]{a} - \sqrt[n]{b}\right) \to \log(a/b)$ that for sufficiently large n we have

$$2^m < 2^n\left(4^{2^{-(n-m-1)}} - 4^{2^{-(n-m)}}\right) < 2^{m+1}$$

for the limit of this expression as $n \to \infty$ is $2^m \log 4$, and $1 < \log 4 < 2$. If now m is any integer and $n > m$, both $n - m$ and $n - m - 1$ are positive, so that we may write the expression above as $2^n(\sqrt{}^{n-m-1}4 - \sqrt{}^{n-m}4)$, the indices of the root signs indicating repetitions. Taking square roots k times, we have

$$2^{m/2^k} < \sqrt{}^k\left(\sqrt{4^n}\left(\sqrt{}^{n-m-1}4 - \sqrt{}^{n-m}4\right)\right) < 2^{(m+1)/2^k}.$$

Now we may take n to be of the form $4(!)^p$ so as to satisfy all the above conditions, and the the expression between the inequality signs will use only four 4's. Since the numbers $2^{m/2^k}$ for integers m and positive integers k are dense in the positive real numbers, we have proved our theorem. (For a negative number we need merely add another – sign.) □

Second perfect number. Sum of the totient functions of the first nine integers.

Theorem 2 If we allow use of the integer part sign, every integer is representable with four fours and every rational number with five.

The first part is obvious, and the second part becomes a corollary of the first when we note that any rational p/q equals $m/4(!)^n$ for suitable integers m and n. □

We may modify Theorems 1 and 2 so as to use other (positive integral) numbers instead of 4's. The only condition is that at most three of these may be 1's.

Finally we pose these questions:

- Is there an "explicit" formula for π with less than seven 4's?
- Is there and explicit formula for γ?
- Are the numbers $\sqrt[n]{(4(!)^m)}$ dense in $x > 1$?

Surface Differences Matter!

Arran Fernandez, Cambridge University

When Grigori Perelman proved the 1904 Poincaré conjecture in 2003, he gave a sketch proof of Thurston's stronger *geometrisation conjecture*, which had been around since 1982. Both concern 3-manifolds. But what about the geometrisation theorem's little sister, which deals with 2-manifolds?

Known as the *classification theorem for closed connected surfaces*, this gives a complete list of all closed connected 2-manifolds up to homeomorphism, enabling any such surface to be slotted into one of two simple categories.

First announced in 1888, with a proof that turned out to be incomplete, the classification theorem was proved in 1907, albeit assuming triangulability, which was only proved in 1925.

Nowadays, it crops up in various second and third-year Tripos courses, including IB Geometry, II Differential Geometry, II Algebraic Topology. But only stated, not proved.

There is an elementary proof, requiring little more than a basic knowledge of triangulations. This was first given by Christopher Zeeman in the 1960s.

Definitions

A **surface**, or **2-manifold**, is a topological space that's locally homeomorphic to $\mathbb{R}^2$. In other words, any point has a small neighbourhood which is approximately flat. Take for example the surface of the Earth: from close up it looks flat, and you need to get a long way away to see that it isn't.

A surface is **connected** if it's all in one piece; and **closed** if it has no boundary and can be expressed as a finite union of discs. So a cylinder isn't closed, and nor is any unbounded surface in 3D, but a sphere and torus are. For brevity, we'll use 'surface' to mean 'closed connected surface'.

Two surfaces are **homeomorphic** if there is a continuous bijection between them with a continuous inverse: intuitively, if they are 'topologically equivalent' in the doughnut-teacup sense.

We assume all surfaces are **triangulable**: in other words, that any surface is topologically equivalent to a polyhedron with flat triangular faces.

The **Euler characteristic** is the quantity $\chi = V - E + F$, where V, E, F are the numbers of vertices, edges, and faces of the triangulation. This is an invariant: any two triangulations of the same surface have the same Euler characteristic.

And off we go...

The Theorem and Proof

The classification theorem states that any closed connected surface S is homeomorphic to one of the following:

- *if it's orientable,* the sphere with g handles glued on, i.e. the g-holed torus, for some $g \geq 0$;
- *if it's non-orientable,* the sphere with h Möbius bands sewn in, for some $h \geq 1$.

Gluing on a handle means removing two small discs on the sphere and sticking the two edges of

Icosahedron and dodecahedron have 30 edges.
The icosidodecahedron has 30 vertices.

a hollow cylinder into the gaps. Sewing in a Möbius band means removing one small disc on the sphere and sticking a Möbius band into the gap. Recall that a Möbius band has only one edge.

We'll need the following two lemmas.

Lemma 1 The Euler characteristic of any surface is at most 2.

Proof: Note that the Euler characteristic of a graph is $\chi = V - E$, since a graph has no faces. If the graph is a tree, i.e. it has no closed loops, then it can be shrunk to a point, so its Euler characteristic is 1. If it isn't a tree, then removing one edge from a closed loop increases the Euler characteristic by 1, and we can get a tree after finitely many such operations. So for a graph we always get $\chi \leq 1$.

Take a triangulation T of a surface S and consider its dual triangulation D, formed by putting a vertex at the centre of each T-face and a face with centre at each T-vertex. Let M be a maximal tree in D, defined as a tree to which no more edges can be added without creating a closed loop, and let $C = D \setminus M$.

Since M is a tree, C is connected. Since M is maximal, M contains all vertices of D. So there are bijections

$$\{T\text{-triangles}\} \leftrightarrow \{M\text{-vertices}\},$$
$$\{C\text{-edges}\} \leftrightarrow \{D\text{-edges}\},$$
$$\{C\text{-vertices}\} \leftrightarrow \{C\text{-vertices}\}.$$

Therefore $\chi(S) = \chi(M) + \chi(C) \leq 2$. □

Lemma 2 If S is a surface which is disconnected by every closed curve on it, then it is homeomorphic to the sphere.

Proof: Let T, D, M, and C be as before. If C contains a loop, then this loop disconnects S; each connected component must contain a D-vertex, and any two D-vertices are joined by edges in M. Contradiction, so C is a tree. Let X be the set of points in S closer to M than to C, and Y be the set of points in S closer to C than to M. Each of X and Y is a fattening up of a tree, so they are both homeomorphic to the disc. But S is just X and Y glued together edge-to-edge, so S is homeomorphic to two discs glued edge-to-edge, i.e. to the sphere. □

We now use the following surgery algorithm on an arbitrary surface S.

1. If S is disconnected by every closed curve on it, stop.
2. If there is a non-disconnecting closed curve on S, remove a thin strip around this curve; this strip must be a cylinder or a Möbius band.
3. If the strip is a cylinder, glue in two discs to the gaps left in S, increasing $\chi(S)$ by 2, and mark both of them with an orientation (clockwise or counterclockwise) so that they agree along the cylinder.
4. If the strip is a Möbius band glue in 1 disc to the gap left in S, increasing $\chi(S)$ by 1.
5. Go to 1.

By Lemma 1, the process stops after finitely many steps. By Lemma 2, the surface we get when it does stop must be a sphere.

Now start from a sphere and reverse the process to get to S in finitely many steps. In each step, we have three possibilities for what needs to be replaced:

a. 1 disc;

b. 2 discs with different orientation (one clockwise, one counterclockwise);

c. 2 discs with the same orientation (both clockwise or both counterclockwise).

If it is (a), we're sewing in one Möbius band. If it is (b), we're gluing in a handle. If it is (c), we're sewing in a Klein bottle. But a Klein bottle is just two Möbius bands, so we can ignore (c) without loss of generality.

So we can get to any surface S by starting with a sphere and putting in finitely many Möbius bands and handles.

If S is orientable, then it can't contain any Möbius strips, so it's a sphere with finitely many handles.

If it is non-orientable, then we must sew in at least one Möbius band. If we ever glue in a handle, then we can transport one of the two differently-oriented discs around this Möbius band so that they've both got the same orientation. This reduces (b) to c), which we've seen reduces to (a). So S is a sphere with finitely many Möbius bands. And this completes the proof. □

Fractals, Compression and Contraction Mapping

Alexander Shannon, Christ's

First published in issue 57, 2005

One of the most often cited applications of the study of fractals is that of their use in image compression. Such an application is not surprising, since seemingly complicated and intricate fractal images have relatively simple mathematical descriptions in terms of iterated mappings. Given also that fractals have been found to model well a wide variety of natural forms, it seems natural that we should try to exploit their self-similar properties to encode images of such forms.

We examine a simple example of a fractal, the Koch curve, to illustrate the principle of encoding a fractal image. Referring to Figure 2, we construct the Koch curve by first taking a line segment of length 1, K_0. We then construct K_1 by combining the images of this segment under four transformations, each involving a dilation of factor 1/3 composed with either or both a rotation and translation. Combining the images of K_1 under the same four transformations yields K_2, and the Koch curve itself (K_∞) is the limit of this process as it is iterated. (Peitgen, in [4], calls this method of drawing fractals the "Multiple Reduction Copy Machine" or MRCM.)

We can see that this object is self-similar, in the sense that we can find arbitrarily small portions of the curve that are related to the whole by a similarity transformation.

The fractal fern of Figure 1 is also the limit of four affine transformations iterated in the same manner. Since each affine transformation may be represented by a 2×2 matrix giving the homogenous part of the transformation and a 2-component vector giving the inhomogeneous (translation) part of the transformation, a figure that is the limit of n iterated affine transformations can be encoded as a collection of $6n$ real numbers – a much more efficient encoding than a pixel-by-pixel representation. These ideas also generalise in an obvious manner to subset of higher dimensional Euclidean space.

We might then ask whether we can measure how 'close' a perfectly self-similar or self-affine fractal is to a given 'imperfect' real life image that we are trying to approximate. We might also ask how many iterations of the kind described above we need to carry out to get a rea-

Figure 1 Real and computer generated Ferns

Freezing point of water at sea level in Fahrenheit. Ninth Happy Humber. $1^1 + 2^2 + 3^3 = 32$.

sonable approximation of the limiting set. More theoretically, we might question whether we can be sure that such iterations will indeed tend to a definite limit, and, given that any such limit will be invariant under the iteration, whether it matters with what set we start. Could we, for example, have begun our construction of the Koch curve with a circle rather than a line segment?

In this article, we shall see that, by considering subsets of Euclidean space as points in a metric space, we can measure how different two images are, and by applying the contraction mapping theorem, we can see that limit sets of the sort described above do exist, that our starting point in their construction does not matter, and we can also obtain an estimate for how rapid the convergence is.

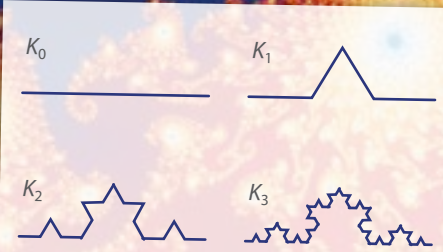


Figure 2 Construction of the Koch curve

Definitions

For reference, we enumerate here a few standard definitions and theorems that we shall use later.

Definition 1 A **metric space** is an ordered pair $(\mathcal{X}, d)$, where $\mathcal{X}$ is a set and $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a function with the following properties:

(i) $d(x,y) \geq 0 \quad \forall\, x,y \in \mathcal{X}$, with $d(x,y) = 0$ if and only if x = y;

(ii) $d(x,y) = d(x,y) \quad \forall\, x,y \in \mathcal{X}$;

(iii) $d(x,z) \leq d(x,y) + d(y,z) \quad \forall\, x,y,z \in \mathcal{X}$.

The notion of convergence of a sequence to a limit carries over to metric spaces in an obvious way, as does the following related notion:

Definition 2 Let (x_n) be a sequence of points in a metric space $(\mathcal{X}, d)$. We say that (x_n) is **Cauchy** if, given $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $m, n \geq N$, $d(x_m, x_n) < \varepsilon$.

Clearly every convergent sequence is a Cauchy sequence. The converse is also true for an important class of metric spaces:

Definition 3 A metric space $(\mathcal{X}, d)$ is **complete** if every Cauchy sequence in $\mathcal{X}$ converges.

We remark that the metric space formed by $\mathbb{R}^n$ with the usual Euclidean metric is complete.

Definition 4 Let $(\mathcal{X}, d)$ be a metric space. Then $f : \mathcal{X} \to \mathcal{X}$ is a **contraction** if there exists a non-negative real number $c < 1$ such that $d(f(x), f(y)) \leq c \times d(x,y)$ for all $x, y \in \mathcal{X}$.

Our central theorem tells us about the behaviour of contractions under iteration (for a proof, see, for example, [3]).

Theorem 5: Contraction Mapping
Let $(\mathcal{X}, d)$ be a non-empty complete metric space and $f : \mathcal{X} \to \mathcal{X}$ a contraction. Then there exists a unique $x_0 \in \mathcal{X}$ such that $f(x_0) = x_0$, and furthermore, $\lim_{n\to\infty} f^n(x) = x_0$ for all $x \in \mathcal{X}$.

In the final section we will refer to a corollary:

Corollary 6 Let $(\mathcal{X}, d)$ be a non-empty complete metric space and $f : \mathcal{X} \to \mathcal{X}$ such that f^n is a contraction. Then the same conclusions hold as for Theorem 5.

For most of the time, we shall restrict our attention to compact subsets of metric spaces.

Definition 7 Let $(\mathcal{X}, d)$ be a metric space. Then we say $A \subseteq \mathcal{X}$ is compact if every covering of A by open sets has a finite subcovering.

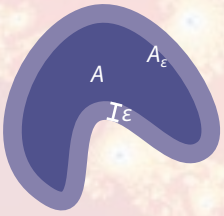
The important properties of compact sets which we need are that they are closed and bounded.

Hausdorff Distance

Our starting point is a way of turning a collection of subsets of Euclidean space into a complete metric space, so that we can talk about limits and convergence, and make use of the considerable information provided by the contraction mapping theorem. The concept we require is due to Hausdorff, who formulated a notion of 'distance' between compact subsets of a metric space which makes the set of compact subsets of a given metric space into a metric space itself. Furthermore, if our initial metric space is complete, then so is the space of compact subsets with the Hausdorff metric.

We require a further concept before introducing the Hausdorff distance itself:

Definition 8 Let A be a subset of a metric space $(\mathcal{X}, d)$. The **ε-collar** of A, denoted A_ε, is the set $\{x \in \mathcal{X} : \exists\, a \in A \text{ with } d(a,x) \le \varepsilon\}$, i.e. the set of all points at a distance at most ε from the set A.



Definition 9 Let A and B be compact subsets of a metric space $(\mathcal{X}, d)$. If we write $\rho'(A,B) = \inf\{\varepsilon > 0 : A \subseteq B_\varepsilon\}$ then the **Hausdorff distance** $\rho(A,B)$ between A and B, is defined by $\rho(A,B) = \max\{\rho'(A,B), \rho'(B,A)\}$.

It follows straightforwardly from the definition that ρ' satisfies all the axioms for a metric space in definition 1 except (ii), so the final part of the definition is essentially a symmetrisation. An alternative definition sometimes used (for example in [3]) but which does the same job is $\rho(A,B) = \rho'(A,B) + \rho'(B,A)$. The proof that the resulting metric space inherits completeness is given in [2] and as an exercise in [3].

The Hutchinson Operator

Now that we have some way of measuring 'closeness' of compact subsets of metric spaces, our next task is to show that the iterated transformation applied in Figure 2 to construct the Koch curve is indeed a contraction, so that we may apply Theorem 5. The following treatment follows quite closely that of [4]. We work in $\mathbb{R}^m$.

We have a collection of affine transformations, $T_1, T_2, \ldots, T_n$, and at each iteration we apply the transformation

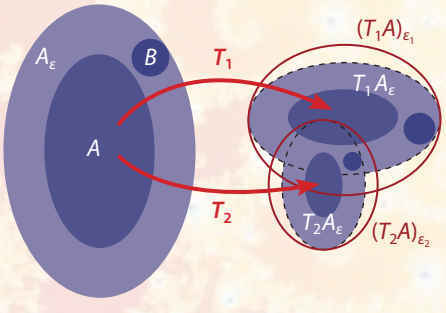
$$T : A \mapsto \bigcup_{i=1}^{n} T_i A.$$

(This is known as the **Hutchinson operator**, after Hutchinson who first analysed its properties.) We impose the condition that each T_i should itself be a contraction with respect to the Euclidean metric, with constant $c_i < 1$.

We now show that T is a contraction with constant $c = \max\{c_1, c_2, \ldots, c_n\}$ on the metric space of compact subsets of $\mathbb{R}^m$ equipped with the Hausdorff metric. (See diagram below for the following.) Let A and B be compact subsets of $\mathbb{R}^m$ with $\rho'(B,A) = \delta$. Then for any $\varepsilon > \delta$ we have $B \subseteq A_\varepsilon$. Clearly then $T_i B \subseteq T_i A_\varepsilon$, for each i, but since T_i is contractive on $\mathbb{R}^m$, $T_i A_\varepsilon \subseteq (T_1 A)_{\varepsilon_i}$, where $\varepsilon_i = c_i\varepsilon < c\varepsilon$. Hence $T_i B \subseteq (T_i A)_{\varepsilon_i} \subseteq (T_1 A)_{c\varepsilon}$, yielding

$$\bigcup_{i=1}^{n} T_i B \subseteq \bigcup_{i=1}^{n} (T_i A)_{c\varepsilon} = \left(\bigcup_{i=1}^{n} T_i A\right)_{c\varepsilon}.$$

So $TB \subseteq (TA)_{c\varepsilon}$ for all $\varepsilon > \delta$, and hence $\rho'(TB, TA) \le c\delta$. Therefore $\rho(TB, TA) \le c \times \rho(A,B)$ and so T is indeed a contraction.



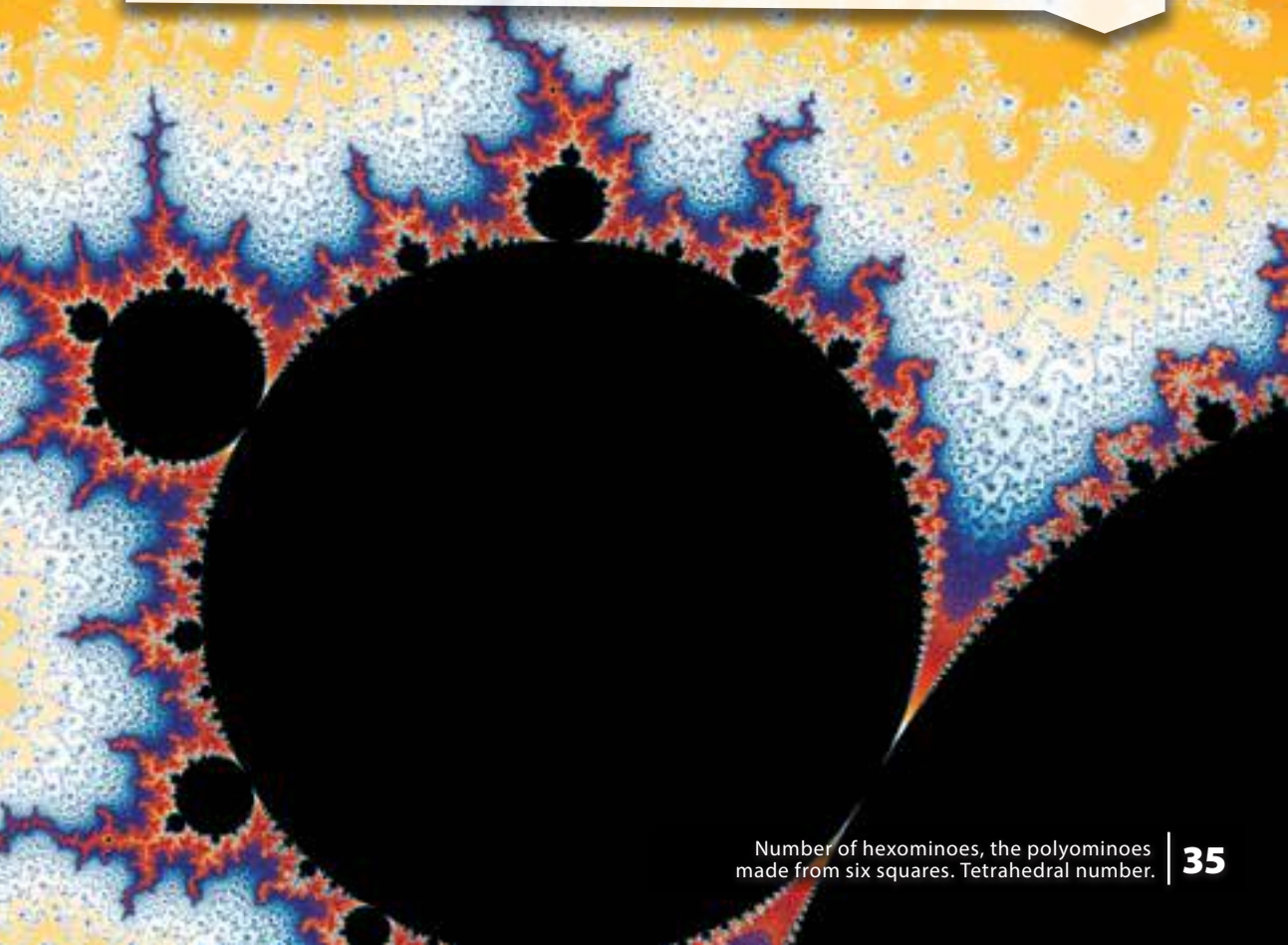
An important practical observation which can be made from the above proof is that the contraction constant calculated for T is equal to the largest of the individual contraction constants of the transformation T_i. It is clear from the proof that, in general, we can do no better than this. In the usual proof of the contraction mapping theorem, it is shown that, for a contraction f with constant c,

$$d\left(f^n(x), f^{n+k}(x)\right) \leq d(x, f(x))\frac{c^n}{1-c}.$$

Since this inequality holds for all k, the expression $c^n/(1 - c)$ provides an estimate for how quickly the iterations converge to the unique fixed point. As might be expected, we see that the larger the constant c, the slower the convergence. Hence the MRCM method of drawing fractals is only as rapid as is allowed by the 'least contractive' contraction. It is, however, worth remarking that a given transformation may or may not be contractive, depending on the choice of metric, and that the contraction constants will vary according to the metric used. Since the notion of Hausdorff distance works for any metric space, not just $\mathbb{R}^m$ with the Euclidean metric, we may certainly replace the Euclidean metric in the above analysis with any other making $\mathbb{R}^m$ into a complete metric space, to be able to draw conclusions about the convergence properties of a wider variety of Hutchinson operators.

Julia Sets

We conclude with some brief, informal remarks about how these ideas may be applied to producing images of another rather famous class of fractals. For a given polynomial $f : \mathbb{C} \to \mathbb{C}$, the **Julia Set** of f, $J(f)$, is the closure of the set of repelling (unstable) fixed and periodic points of f. This is non-trivially equivalent to the definition as the boundary of the basins of attraction of the attractive fixed points of f (for details see [1]), and the set $J(f)$ has the property that $f(J) = f^{-1}(J) = J$. The most famous example of these objects are those associated with the mapping $f : z \mapsto z^2 + c$ for various $c \in \mathbb{C}$ (like the example shown in Figure 3). In this case we notice that the inverse mapping



$f^{-1} : z \mapsto \{\pm\sqrt{z-c}\}$ seems to play the role of a non-linear Hutchinson operator, in that each point (other than c itself) has two images, and the fractal of interest is invariant under the transformation.

We might well then ask whether the mapping is contractive. Here a partial answer is suggested by the theory of conformal mappings, which tells us that for a conformal mapping $g : \mathbb{C} \to \mathbb{C}$, the approximate scaling in length near a point z_0 in $\mathbb{C}$ is $|g'(z_0)|$. The criterion for a fixed point z_0 of a mapping g to be attractive, viz. $|g'(z_0)| < 1$, is therefore the same as the criterion for the mapping to be locally contractive. Any point close to $J(f)$ is, by definition, close to some repelling periodic point of f (whose period we shall denote by p), which in turn will be an attractive periodic point of $f_1^{-1} : z \mapsto +\sqrt{z-c}$ and $f_2^{-1} : z \mapsto -\sqrt{z-c}$. Hence the iterate T^p of the Hutchinson operator T defined by these two mappings will be a local contraction, and so Corollary 6 suggests that, at least if we consider sets not 'too far' in terms of Hausdorff distance from $J(f)$, the iteration will converge in the same manner as for the self-affine fractals discussed above. In fact the convergence is very good, and although after a finite time the iterates do not in general approximate all parts of the Julia set evenly, this is how many fractal packages produce their images of Julia Sets.

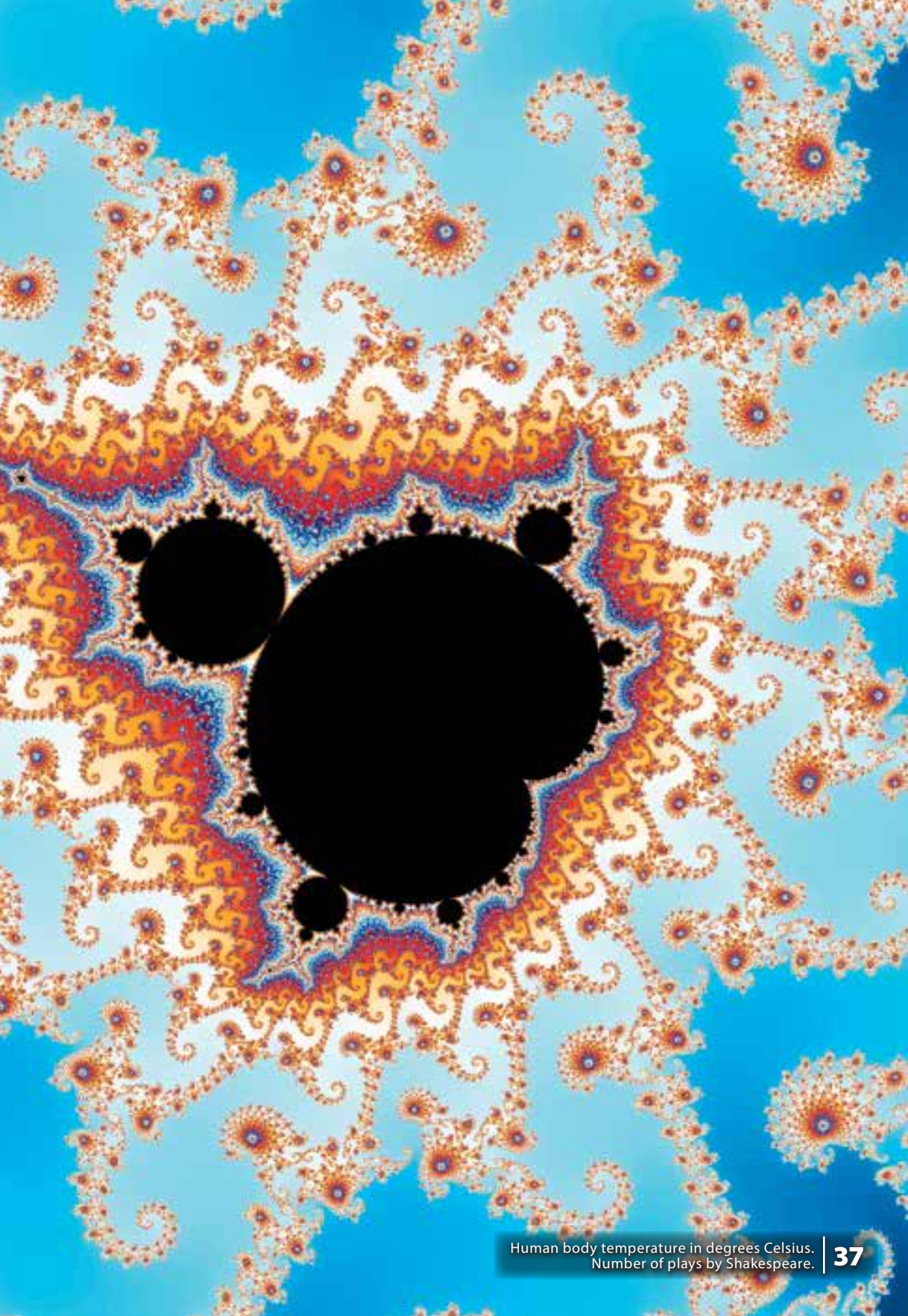
Figure 3 A Julia Set

References

1. K. J. Falconer, *Fractal Geometry*, Wily (1990)
2. F. Hausdorff, *Set Theory*, Chelsea Pub. Co. (1962)
3. T. W. Körner, *A Companion to Analysis*, Americal Mathematical Society (2003)
4. H.-O. Peitgen, H. Jürgens and D. Saupe, *Chaos and Fractals*, Springer Verlag (1992)

Images of the Mandelbrot Set

Smallest non-trivial square-triangular number. The sum of the first 36 integers is 666.



Human body temperature in degrees Celsius.
Number of plays by Shakespeare.

Stein's Paradox

Dr Richard J. Samworth, Statslab Cambridge

Perhaps the most surprising result in Statistics arises in a remarkably simple estimation problem. Let $X_1, \ldots, X_p$ be independent random variables, with $X_i \sim N(\theta_i, 1)$ for $i = 1, \ldots, p$. Writing $X = (X_1, \ldots, X_p)^T$, suppose we want to find a good estimator $\hat{\theta} = \hat{\theta}(X)$ of $\theta = (\theta_1, \ldots, \theta_p)^T$. To define more precisely what is meant by a *good* estimator, we use the language of statistical decision theory. We introduce a **loss function** $L(\hat{\theta}, \theta)$, which measures the loss incurred when the true value of our unknown parameter is θ, and we estimate it by $\hat{\theta}$. We will be particularly interested in the squared error loss function $L(\hat{\theta}, \theta) = \|\hat{\theta} - \theta\|^2$, where $\|\cdot\|$ denotes the Euclidean norm, but other choices, such as the absolute error loss $L(\hat{\theta}, \theta) = \sum_{i=1}^{p} |\hat{\theta}_i - \theta_i|$ are of course perfectly possible.

Now $L(\hat{\theta}, \theta)$ is a random quantity, which is not ideal for comparing the overall performance of two different estimators (as opposed to the losses they each incur on a particular data set). We therefore introduce the **risk function**

$$R(\hat{\theta}, \theta) = \mathbb{E}\{L(\hat{\theta}, \theta)\}.$$

If $\hat{\theta}$ and $\tilde{\theta}$ are both estimators of θ, we say $\hat{\theta}$ **strictly dominates** $\tilde{\theta}$ if $R(\hat{\theta}, \theta) \leq R(\tilde{\theta}, \theta)$ for all θ, with strict inequality for some value of θ. In this case, we say $\tilde{\theta}$ is **inadmissible**. If $\hat{\theta}$ is not strictly dominated by any estimator of θ, it is said to be **admissible**. Notice that admissible estimators are not necessarily sensible: for instance, in our problem above with $p = 1$ and the squared error loss function, the estimator $\hat{\theta} = 37$ (which ignores the data!) is admissible. On the other hand, decision theory dictates that inadmissible estimators can be discarded, and that we should restrict our choice of estimator to the set of admissible ones.

This discussion may seem like overkill in this simple problem, because there is a very obvious estimator of θ: since all the components of X are independent, and $\mathbb{E}(X_i) = \theta_i$ (in other words X_i is an **unbiased** estimator of θ_i), why not just use $\hat{\theta}^0(X) = X$? Indeed, this estimator appears to have several desirable properties (for example, it is the maximum likelihood estimator and the uniform minimum variance unbiased estimator), and by the early 1950's, three proofs had emerged to show that $\hat{\theta}^0$ is admissible for squared error loss when $p = 1$. Nevertheless, Stein (1956) stunned the statistical world when he proved that, although $\hat{\theta}^0$ is admissible for squared error loss when $p = 2$, it is inadmissible when $p \geq 3$. In fact, James and Stein (1961) showed that the estimator

$$\hat{\theta}^{JS}(X) = \left(1 - \frac{p-2}{\|X\|^2}\right) X$$

strictly dominates $\hat{\theta}^0$. The proof of this remarkable fact is relatively straightforward, and is given in the Appendix.

Largest even number which cannot be written as the sum of two odd composite numbers.

One of the things that is so surprising about this result is that even though all of the components of X are independent, the ith component of $\hat{\theta}^{JS}$ depends on all of the components of X. To give an unusual example to emphasise the point, suppose that we were interested in estimating the proportion of the US electorate who will vote for Barack Obama, the proportion of babies born in China that are girls and the proportion of Britons with light-coloured eyes. Then our James–Stein estimate of the proportion of democratic voters depends on our hospital and eye colour data! The reader might reasonably complain that in the above examples, the data would be binomially rather than normally distributed. However, one can easily transform binomially distributed data so that it is well approximated by a normal distribution with unit variance (see the baseball example below), and then consider the estimation problem on the transformed scale, before applying the inverse transform.

Geometrically, the James–Stein estimator shrinks each component of X towards the origin, and it is therefore not particularly surprising that the biggest improvement in risk over $\hat{\theta}^0$ comes when $\|\theta\|$ is close to zero; see Figure 1 for plots of the risk functions of $\hat{\theta}^0$ and $\hat{\theta}^{JS}$ when $p = 5$. A simple calculation shows that $R(\hat{\theta}^{JS}, 0) = 2$ for all $p \geq 2$, so the improvement in risk can be substantial when p is moderate or large. In terms of choosing a point to shrink towards, though, there is nothing special about the origin, and we could equally well shrink towards any pre-chosen $\theta_0 \in \mathbb{R}^p$ using the estimator

$$\hat{\theta}^{JS}_{\theta_0}(X) = \theta_0 + \left(1 - \frac{p-2}{\|X-\theta_0\|^2}\right)(X-\theta_0).$$

In this case, we have $R(\hat{\theta}^{JS}_{\theta_0}, \theta - \theta_0) = R(\hat{\theta}^{JS}, \theta)$, so $\hat{\theta}^{JS}_{\theta_0}$ still strictly dominates $\hat{\theta}^0$ when $p \geq 3$.

Note that the shrinkage factor in $\hat{\theta}^{JS}_{\theta_0}$ becomes negative when $\|X - \theta_0\|^2 < p - 2$, and indeed it can be proved that $\hat{\theta}^{JS}_{\theta_0}$ is strictly dominated by the positive-part James–Stein estimator

$$\hat{\theta}^{JS}_{+,\theta_0}(X) = \theta_0 + \left(1 - \frac{p-2}{\|X-\theta_0\|^2}\right)_+ (X-\theta_0),$$

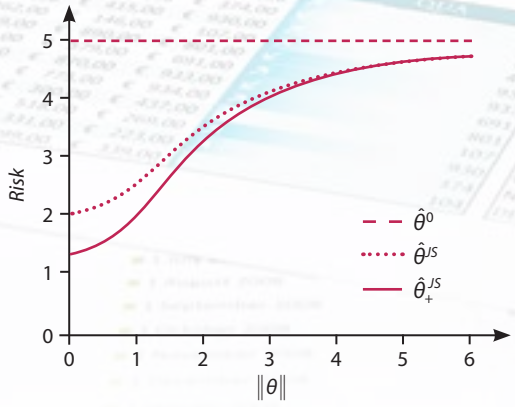
where $x_+ = \max(x, 0)$. The risk of the positive-part James–Stein estimator $\hat{\theta}^{JS}_+ = \hat{\theta}^{JS}_{+,0}$ is also included in Figure 1 for comparison. Remarkably, even the positive-part James–Stein estimator is inadmissible, though it cannot be improved by much, and it took until Shao and Strawderman (1994) to find a (still inadmissible!) estimator to strictly dominate it.

Generalisations and Related Problems

It is natural to ask how crucial the normality and squared error loss assumptions are to the Stein phenomenon. As a consequence of many papers written since Stein's original masterpiece, it is now known that the normality assumption is not critical at all; similar (but more complicated) results can be proved for very wide classes of distributions. The original result can also be generalised to different loss functions, but there is an important caveat here: the Stein phenomenon only holds when we are interested in simultaneous estimation of all components of θ. If our loss function were $L(\hat{\theta}, \theta) = (\hat{\theta}_1 - \theta_1)^2$, for example, then we could not improve on $\hat{\theta}^0$. This explains why it wouldn't make much sense to use the James–Stein estimator in our bizarre example above; it is inconceivable that we would be simultaneously interested in three such different quantities to the extent that we would want to incorporate all three estimation errors into our loss function.

Although Stein's result is very clean to state and prove, it may seem somewhat removed from practical statistical problems. Nevertheless, the idea at the heart of Stein's proposal, namely that of employing shrinkage to reduce variance (at the expense of introducing bias) turns out to be a very powerful one that has had a huge impact on statistical methodology. In particular, many modern statistical models may involve thousands or even millions of parameters (e.g. in microarray experiments in genetics, or fMRI studies in neuroimaging); in such circumstances, we would almost certainly want estimators to set some of the parameters to zero, not only to improve performance but also to ensure the interpretability of the fitted model.

Sum of five consecutive primes (3 + 5 + 7 + 11 + 13) and the first three powers of three (3 + 9 + 27).



▲ Figure 1: Risks with respect to squared error loss of the usual estimator $\hat{\theta}^0$, the James–Stein estimator $\hat{\theta}^{JS}$ and the positive-part James–Stein estimator $\hat{\theta}_+^{JS}$ when $p = 5$.

Player	n_i	Z_i	π_i
Baines	415	0.284	0.289
Barfield	476	0.246	0.256
Bell	583	0.254	0.265
Biggio	555	0.276	0.287
Bonds	519	0.301	0.297
Bonilla	625	0.280	0.279
Brett	544	0.329	0.305
Brooks Jr.	568	0.266	0.269
Browne	513	0.267	0.271

▲ Table 1: Table showing number of times at bat n_i, batting average Z_i in 1990, and career batting average π_i, of $p = 9$ baseball players.

Another important problem that is closely related to estimation is that of constructing a confidence set for θ, the aim being to give an idea of the uncertainty in our estimate of θ. Given $\alpha \in (0,1)$, an **exact $(1-\alpha)$-level confidence set** is a subset $C = C(X)$ of $\mathbb{R}^p$ such that, whatever the true value of θ, the confidence set contains it with probability exactly $1-\alpha$. The usual, exact $(1-\alpha)$-level confidence set for θ in our original normal distribution set-up is a sphere centred at X. More precisely, it is

$$C^0(X) = \{\vartheta \in \mathbb{R}^p : \|\vartheta - X\|^2 \le \chi^2_p(\alpha)\},$$

where $\chi^2_p(\alpha)$ denotes the upper α-point of the χ^2_p distribution (in other words, if $Z \sim \chi^2_p$, then $\mathbb{P}\{Z > \chi^2_p(\alpha)\} = \alpha$). But in the light of what we have seen in the estimation problem, it is natural to consider confidence sets that are spheres centred at $\hat{\theta}_+^{JS}$ (or $\hat{\theta}^{JS}_{+,\theta_0}$, for some $\theta_0 \in \mathbb{R}^p$). Since the distribution of $\|\hat{\theta}_+^{JS} - \theta\|^2$ depends on $\|\theta\|$, we can no longer obtain an exact $(1-\alpha)$-level confidence set, but it may be possible to construct much smaller confidence sets – using bootstrap methods to obtain the radius, for example – which still have at least $(1-\alpha)$-level coverage (e.g. Samworth, 2005).

A baseball data example

The following example is adapted from Samworth (2005). The data in Table 1 give the baseball batting averages (number of hits divided by number of times at bat) of $p = 9$ baseball players, all of whom were active in 1990. The source was *www.baseball-reference.com*. For $i = 1, \ldots, p$, let n_i and Z_i respectively denote the number of times at bat and batting average of the ith player during the 1990 season. Further, let π_i denote the player's true batting average, taken to be his career batting average. (Each player had at least 3000 at bats in his career.) We consider the model where $Z_1, \ldots, Z_p$ are independent, with $Z_i \sim n_i^{-1}\,\mathrm{Bin}(n_i, \pi_i)$.

We make the transformation

$$X_i = \sqrt{n_i}\,\sin^{-1}(2Z_i - 1),$$

and let $\theta_i = \sqrt{n_i}\,\sin^{-1}(2\pi_i - 1)$, which means that X_i is approximately distributed as $N(\theta_i, 1)$. A heuristic argument (which can be made rigorous) to justify this is that by a Taylor expansion applied to the function $g(x) = \sqrt{n_i}\,\sin^{-1}(2x-1)$, we have

$$X_i - \theta_i = g(Z_i) - g(\pi_i) \approx g'(\pi_i)(Z_i - \pi_i) = \frac{\sqrt{n_i}(Z_i - \pi_i)}{\sqrt{\pi_i(1-\pi_i)}},$$

and this latter expression has an approximate $N(0,1)$ distribution when n_i is large, by the central limit theorem. In fact, since $\min_i n_i \ge 400$, an exact calculation gives that the variance of each X_i is between 1 and 1.005 for $\pi_i \in [0.2, 0.8]$. For our prior guess $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,p})^{\mathrm{T}}$, we take $\theta_{0,i} = \sqrt{\bar{n}}\,\sin^{-1}(2\pi_0 - 1)$, with $\pi_0 = 0.275$ and $\bar{n} = p^{-1}\sum_{i=1}^p n_i$. We find that $\|X - \theta\|^2 = 2.56$, somewhat below its expected value of around 9, though since the variance of a χ^2_9 random variable is 18, this observation is only around 1.5 standard deviations away from its mean. On the other hand, $\|\hat{\theta}^{JS}_{+,\theta_0} - \theta\|^2 = 1.50$, so Stein estimation does provide an improvement in this case.

Only number whose letters are in alphabetical order. Venus returns to the same point in the night sky every 40 years.

Letting $\pi = (\pi_1, \ldots, \pi_p)$ and recalling that θ is a function of π, the usual 95% confidence set for π is

$$\{\pi \in [0,1]^p : \|X - \theta\|^2 \leq 16.9\}.$$

On the other hand, the 95% confidence set for π constructed using the bootstrap approach is

$$\{\pi \in [0,1]^p : \|\hat{\theta}^{JS}_{+,\theta_0}(X) - \theta\|^2 \leq 12.5\}.$$

Numerical integration gives that the volume ratio of the bootstrap confidence set to the usual confidence set in this case is 0.26, so the benefits of having centred the confidence set more appropriately are quite substantial.

References

1. W. James, and C. Stein, *Estimation with quadratic loss*, Proc. Fourth Berkeley Symposium, 1, 361–380, Univ. California Press (1961)
2. R. Samworth, *Small confidence sets for the mean of a spherically symmetric distribution*, J. Roy. Statist. Soc., Ser. B, 67, 343–361 (2005)
3. C. Stein, *Inadmissibility of the usual estimator of the mean of a multivariate normal distribution*, Proc. Third Berkeley Symposium, 1, 197–206, Univ. California Press (1956)
4. P. Y.-S. Shao and W. E. Strawderman, *Improving on the James–Stein positive-part estimator*, Ann. Statist., 22, 1517–1538 (1994)

Appendix

First note that since $\|X-\theta\|^2 \sim \chi^2_p$, we have $R(\hat{\theta}^0, \theta) = p$ for all $\theta \in \mathbb{R}^p$. To compute the risk of the James–Stein estimator, note that we can write

$$\begin{aligned} R(\hat{\theta}^{JS}, \theta) &= \mathbb{E}\left\{\left\| X - \theta - \frac{(p-2)X}{\|X\|^2}\right\|^2\right\} \\ &= p - 2(p-2)\sum_{i=1}^{p}\mathbb{E}\left\{\frac{X_i(X_i-\theta_i)}{\|X\|^2}\right\} + (p-2)^2\mathbb{E}\left(\frac{1}{\|X\|^2}\right). \end{aligned}$$

Consider the expectation inside the sum when $i = 1$. We can simplify this expectation by writing it out as an n-fold integral, and computing the inner integral by parts:

$$\begin{aligned} \mathbb{E}\left\{\frac{X_1(X_1-\theta_1)}{\|X\|^2}\right\} &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{x_1}{\|x\|^2}\times\frac{(x_i-\theta_i)}{(2\pi)^{p/2}}e^{-\|x-\theta\|^2/2}\,dx_1\ldots dx_p \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\|x\|^2-2x_1^2}{\|x\|^4}\times\frac{1}{(2\pi)^{p/2}}e^{-\|x-\theta\|^2/2}\,dx_1\ldots dx_p, \end{aligned}$$

since the integrated term vanishes. Repeating virtually the same calculation for components $i = 2, \ldots, p$, we obtain

$$\begin{aligned} \sum_{i=1}^{p}\mathbb{E}\left\{\frac{X_i(X_i-\theta_i)}{\|X\|^2}\right\} &= \sum_{i=1}^{p}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{\|x\|^2-2x_i^2}{\|x\|^4}\right)\frac{1}{(2\pi)^{p/2}}e^{-\|x-\theta\|^2/2}\,dx_1\ldots dx_p \\ &= \sum_{i=1}^{p}\mathbb{E}\left(\frac{\|X\|^2-2X_i^2}{\|X\|^4}\right) = (p-2)\mathbb{E}\left(\frac{1}{\|X\|^2}\right). \end{aligned}$$

We therefore conclude that

$$R(\hat{\theta}^{JS}, \theta) = p - (p-2)\mathbb{E}\left(\frac{1}{\|X\|^2}\right) < p$$

for all $\theta \in \mathbb{R}^p$, as required.

Optimal Card Shuffling

Martin Mellish

First published in issue 36, 1973

One shuffles a pack of cards by dividing it into two portions and merging them onto a flat surface by riffling with the fingertips. The problem that concerns us is: *How many times do we have to do this before the pack is perfectly shuffled, and what exactly is the best method of shuffling?* A pack of cards is said to be perfectly shuffled if

(i) either all possible decks are equally likely (this is what we require for patience),

(ii) or all possible deals of the deck into hands are equally likely (this is what we require for Bridge).

The objective of this article is to find an explicit solution of the problem in case (i) and give some guidelines in case (ii) which may enable some interested readers to solve the problem in that case also. We will start by proving some subsidiary results, after noting that any sequence of shuffles of a pack of N cards can be regarded as a random permutation of the integers from 1 to N.

Lemma 1 If $D(i)$ for $i = 1, \ldots, k$ is the set of decks which can arise from shuffling a given pack once, and if there exists $x_i \geq 0$ such that $\sum_{i=1}^{k} x_i = 1$, then there exists a shuffling strategy under which $\mathbb{P}(D(i)) = x_i$.

Proof: Define $P_{j,L}(\alpha) = \mathbb{P}(\alpha, L \mid \alpha)$, where α is a sequence of the form (L, R, R, L, R, L, …) telling us whether each of the first $j-1$ cards fell from the right or the left. Then the shuffling strategy is defined by the $P_{j,L}(\alpha)$. Set $S(\alpha)$ be the subset of $D(i)$ which corresponds to shuffles beginning with α (the reader should satisfy himself that the obvious correspondence between shuffles and decks is a bijection). Then we set

$$P_{j,L}(\alpha) = \frac{\sum_{i:D(i)\in S(\alpha,L)} x_i}{\sum_{i:D(i)\in S(\alpha)} x_i}.$$

A simple inductive check shows that this is the required strategy. □

Lemma 2 There is a bijection between decks obtained by shuffling a new pack of N cards m times or less, and sequences a_i for $i = 1, \ldots, N$ satisfying

(i) $1 \leq a_i \leq N$;

(ii) $a_i \neq a_j$ whenever $i \neq j$;

(iii) a_i can be expressed as the union of p subsequences $b^k_{L_i}$, all of which satisfy $b^k_{L_i} - b^k_{L_{i-1}} = 1$, where $p \leq 2^m$.

Proof: Label each card of the new deck with an integer from 1 to N, starting from the top and working down. Then any deck after m shuffles can be represented as a sequence satisfying (i) and (ii). Furthermore, the 2^m subsets of the deck, such that two elements of the same set were in the same portion of the pack after every cut (some sets of which may be empty), correspond to the subsequences $b^k_{L_i}$, for a shuffle cannot change the order of such a subset.

Conversely, suppose we are given the subsequence $b^k_{L_i}$. We may assume without loss of generality that

each sequence $b^k_{L_i}$ is the maximal subsequence satisfying (iii) and containing $b^k_{L_i}$, otherwise one would concatenate subsequences until this were true.

Define a vector $\mathbf{v}(k)$ with $\lceil \log_2 p \rceil$ components, such that $v_i(k)$ is the coefficient of 2^{i-1} in the binary expansion of $k-1$. Assume also without loss of generality that $b^k_{L_i}$ is the least number not equal to one of $\bigcup_{j=1}^{k-1} b^j_{L_i}$ by ordering the subsequences in a canonical manner.

Then corresponding to this sequence a_i we define a sequence $\lceil \log_2 p \rceil$ shuffles. The sequences $b^k_{L_i}$ are all in the top half at the j'th cut if $v_{N\,1-j}(k) > 0$ and are all in the bottom half otherwise. A simple inductive argument (left to the reader) shows that such a sequence of shuffles exists and is unique. Furthermore, since p is not greater than 2^m, $\lceil \log_2 p \rceil$ it is not greater than m. □

We now proceed to define a sequence of optimal shuffling procedures P_1, P_2, …, P_n such that, if $\{D_n(i),\ i = 1, \ldots, k_n\}$ is the set of all decks which can be obtained by shuffling a new deck n times or less, then all the $D_n(i)$ are equally likely after the successive application of P_1, …, P_n. We define P_n inductively: let P_1, …, P_{n-2} be perfect shuffles and let $M_n(j)$ be the number of elements of $D_{n-1}(i)$ for $i = 1, \ldots, k_{n-1}$, which can give rise to $D_n(i)$ after a single cut and shuffle. Let $\{D_n(k_{C,j}),\ i = 1, \ldots, L_{n,C}\}$ be the set of all decks obtainable by a sequence of n shuffles in which C was the last cut. The we define the probability that the cut in P_n is C to be

$$\sum_{j=1}^{L_{n,C}} \frac{1}{M_n(j)}.$$

By applying Lemma 1 we can construct a shuffling strategy such that

$$\mathbb{P}(\text{cut} = C, \text{deck} = D_n(k_{C,j})) = \frac{1}{k_n\ M_n(k_{C,j})}.$$

Summing over all possible cuts we get $\mathbb{P}(\text{deck} = D_n(k_{C,j})) = 1/k_n$. Thus we have defined P_n. The reader might care, as an instructive exercise, to work out P_1, the optimal procedure for the first shuffle.

It is now easy to see that the smallest number of shuffles necessary to randomise a pack of N cards completely is $\lceil \log_2 N \rceil$: we consider the deck in which the original order of the pack is reversed and apply Lemma 2 to see that $D_n(i)$, case $P_1\ P_2 \ldots P_n$, will perfectly shuffle the pack. This is a nice result, and what one would expect from information theory.

The corresponding result for case (ii) mentioned in the introductory paragraph would be that one requires $\lceil \log_2 M \rceil$ shuffles, where M is the number of players. Unfortunately the beauty of this result is spoiled by the fact that it is false – the true value is $\lceil \log_2 f(M,N) \rceil$, where $f(M,N)$ is the least f such that for any M hands of N/M cards H^i_j for $j = 1, \ldots, N/M$ and $i = 1, \ldots, M$, there exists a sequence a_i of N integers as defined in Lemma 2, with the number of subsequences b^k less than or equal to $f(M,N)$ and also

$$\bigcup_{j=1}^{N/M} H^i_j = \bigcup_{j=0}^{N-1} a_{j+j_M}.$$

Unfortunately it is not clear how $f(M,N)$ can be evaluated; all that is clear is that $M \le f(M,N) \le N$. Perhaps one of our readers would care to earn himself a place in the hearts of Bridge players everywhere by solving this problem?

Harmonic Game Theory

Blanche Descartes

First published in issue 16, 1953

Most mathematicians know the theory of the game of Nim, described in books on mathematical recreations. But few seems to be aware of Dr P. M. Grundy's remarkable generalisation, published in Eureka 2 in 1939. Consider a game Γ in which 2 players move alternately, and the last player wins (moving to a "terminal position"). Define inductively a function $G(P)$ of the position P as follows:

(a) if P is terminal, $G(P) = 0$;

(b) if there are permitted moves from P to Q, from P to R, from P to S, and so on, then $G(P)$ is the least non-negative integer different from all of $G(Q)$, $G(R)$, $G(S)$, …

It follows that if $0 \le r < G(P)$ there is a move from P to some R with $G(R) = r$, but no move to any position U with $G(U) = G(P)$. If positions P with $G(P) = 0$ are called "safe," the winning strategy is to move always to a safe position: either this is terminal, and wins immediately, or the opponent moves to an unsafe position and the cycle repeats.

Now imagine the players engaging in a "simultaneous display" of k games $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$ of this sort, the rule being that each player in turn makes a move in one and only one game, or if he cannot move in any game he loses. Let $P_1, P_2, \ldots, P_k$ be the positions in the respective games $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$. Then Grundy's Theorem states that

(i) this combined position is safe if and only if k heaps of $G(P_1), G(P_2), \ldots, G(P_k)$ counters respectively form a safe combination in Nim,

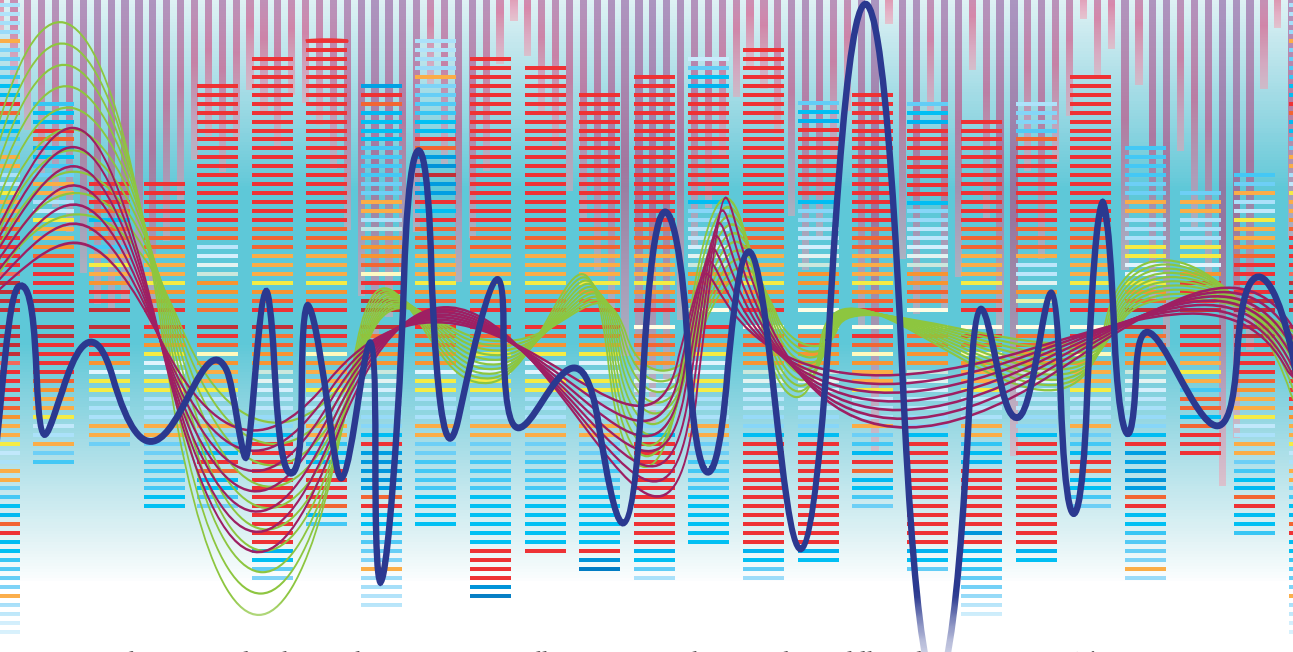
(ii) more generally, the G function of the combined position is the "nim-sum" of the separate $G(P_s)$, i.e. obtained by writing the $G(P_s)$ in the scale of 2 and adding columns mod 2.

For no player can gain any advantage by moving so as to increase any $G(P_s)$, as the opponent can restore the *status quo*. And if only decreases in $G(P_s)$ are considered, the game is identical with Nim, thus proving assertion (i). Therefore $G(P) = g$ if and only if the combined position (P, P') is safe, where $G(P') = g$. From that (ii) follows fairly readily.

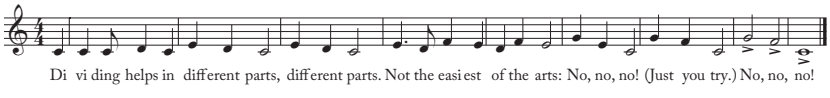
It follows that we can analyse any such combined game completely, provided that we can find the $G(P_s)$ for the component positions. Nim is an example; a heap H_x of x counters constitutes a component position, since each player in turn alters one heap only, and $G(H_x) = x$. Many variants of Nim are similarly analysed. Less trivial is Grundy's game, in which any one heap is divided into two unequal (non-empty) parts. Thus heaps of 1, 2, are terminal, with $G = 0$, a heap of 3 can only be divided into 2 + 1, which is terminal, so $G(H_3) = 1$. Generally $G(H_x)$ in Grundy's game is the least integer > 0 different from all nim-sums of $G(H_y)$ and $G(H_{x-y})$ for $0 < \frac{1}{2}x$. The series goes

x	=	0	1	2	3	4	5	6	7	8	9
$G(H_x)$	=	0	0	0	1	0	2	1	0	2	1

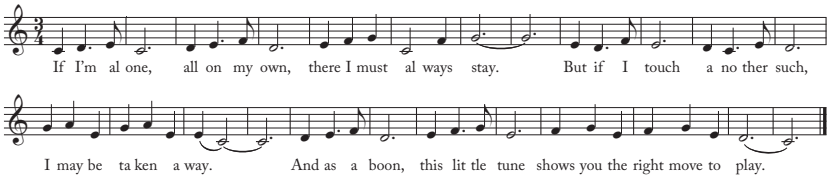
continuing with 0, 2, 13, 2, 1, 3, 2, 4, 3, 0, 4, 3, 0, 4, 3, 0, 4, 1, 2, 3, 1, 2, 4, 1, 2, 4, 1, 2, … This curious "somewhat periodic series" seems to be trying to



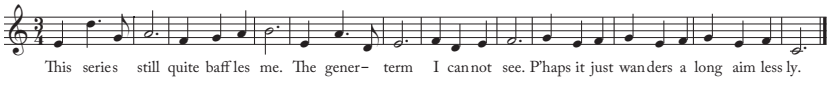
have period 3, but with jumps continually occurring. Richard Guy confirmed this up to $x = 300$. He suggested that it might be played on a piano, taking 0 to be middle C, 1 = D, 2 = E, etc. The inner meaning then became evident:



Guy also worked with rows R_x of x counters, in which certain sets of consecutive counters could be extracted (thus possibly leaving two shorter rows, one each side of the extracted set). In his "**·6**" game, any one counter can be removed, except an R_1 (= a single counter standing on its own). The $G(R_x)$ series ($x = 1, 2, \ldots$) is a waltz. (Note that some notes span two bars.)



But at this point the tune completely broke down. I asked Guy if he could think of any reason for that. He said, "Yes, an error I made in the calculation." After correction the waltz proceeds:



This tries to be periodic with period 26, but jumps keep appearing. Many other such games give tuneful, somewhat periodic series, for no evident reason. Guy discovered two curious exceptions: his "**·4**", remove 1 counter not at the end of a row, has exact period 34 for $x \geq 54$, and Kayles, remove 1 or 2 adjacent counters, has exact period 12 for $x \geq 71$. Thus these games have a complete analysis. But generally it might be helpful to bring in a professional musician to study number theory. Perhaps a thorough study of the Riemann Hypothesis will uncover the Lost Chord. After all, why not?

The Logic of Logic

Zoe Wyatt, Newnham

In the early 20th century mathematicians embarked on a quest to find a secure foundation for their subject, based on the use of axioms and rigorous logic.

By itself however, a collection of axioms is not very useful, since it does not generate anything on its own. Only in conjunction with some logic, the rule of inference for example, do axioms lead to results.

The rule of inference, or *modus ponens*, says:

$$\frac{P \Rightarrow Q,\ P}{\therefore Q}$$

In this and any mathematical expression, we use symbols to express ourselves. As with words in language and expression in conversation, we rely on 'tools' to convey the substance of our thoughts. Of course with words we can find inconsistencies:

I fit into my shirt.

My shirt fits into my bag.

Therefore I fit into my bag.

Though trivial, this shows that words have an underlying associated meaning. With this restriction in mind, we could fix the above by replacing 'bag' with 'very large wardrobe'.

Mathematics avoids such a restriction altogether; swapping P with Q in the *modus ponens* would still yield the same results. Obvious you might think, but philosophically this structural difference is of great importance.

Hilbert's Finitism

Start with the statement: $\forall\ x \in \mathbb{Z}$, $\phi(x)$ is true, where $\phi(x)$ can be precisely one of 'true' or 'false'. If we negate this statement, would you imagine checking an infinite number of x's for falsity in ϕ? Or perhaps spot a suspicious looking x and prove him to be a counterexample?

In the early 1920s, Hilbert was losing sleep on such matters. Or to be precise, he was concerned with making meaningful propositions and methods of reasoning which did not require the acceptance of infinite entities. This finitary viewpoint is particularly important in the context of mathematical operations, by only allowing arguments which can be translated into a finite set of propositions starting from a finite set of axioms.

Of particular concern was the Quantifier Law of Excluded Middle (QLEM):

Every x satisfies ϕ, or some x satisfies the negation of ϕ,

where ϕ is again a statement which is either true or false.

Hilbert held a finitary view, meaning that if the domain being tested was infinite, the QLEM was not to be trusted. How could he know the value of φ (x) for any one of an infinite of x's? More generally, the finite belief prevented simultaneously allowing a property to be associated with infinitely many objects. In our case, it means we cannot apply an infinite conjunction to the integers:

$$\phi(1) \text{ and } \phi(2) \text{ and } \phi(3) \ldots$$

Number of human Chromosomes.
Centred triangular number. Erdős–Woods number.

If instead we are equipped with only finite procedures, then given a particular integer, we are able to prove the instance of $\phi(n)$ for precisely that case. Hence negating this statement is not universal. So *not* $\phi(n)$ indicates that every instance of 'n fails property ϕ' is true, but it does not tell us that not every instance of $\phi(n)$ is true. Put more simply, the following statements are not *universally* logically equivalent:

- Not every x satisfies $\phi(x)$;
- Some x satisfies not-$\phi(x)$.

Thus QLEM fails. More generally though, this shows that problems in our base assumptions need to be addressed to prevent ramifications further on.

Truth and its Limitations

Also during the late 19th and 20th Centuries, many mathematicians began to question the limits on what kind of mathematical objects could be represented and manipulated. Aside from the popular Gödel's Theorem of Incompleteness, and Russell's Paradox, a key if slightly less popular result is Tarski's Undefinability Theorem. Published in 1936, this (very informally) states:

> *Given a formal arithmetic, a true arithmetical statement cannot be defined in that arithmetic.*

To explain the original, technical form of this theorem would take too long here [2], however we see from the above statement of the theorem that formal languages containing semantic terms like "true", will always give a paradox when these terms are self-referenced. Tarski addressed this by making the distinction between semantically closed and semantically open languages. He defined a semantically closed language to be one in which it is possible for a single sentence to predicate (determine) truth or falsehood in another sentence in the same language, or even of itself. Put simply, a semantically closed language can apply semantic properties to the terms that express semantic properties.

This suggests that for a semantically open language to achieve consistency, we need to use a more powerful language, called a metalanguage, in order to be able to define a truth predicate. One of the most common uses of metalanguages is in computer science, such as the Backus-Naur Form developed in the 1960s, to describe the syntax of computer programming languages.


David Hilbert (1862 – 1943)

So Have We Done Anything?

Understanding what makes a statement true or false, and how our mathematics relates to our thoughts, has many times uncovered the limitations of underlying assumptions. Such questioning often leads to fruitful ways of new thinking, a key example being the development of hyperbolic geometry in the 19th Century by the rejection of Euclid's 2000 year old parallel line axiom.

Similarly the work of Hilbert, Tarski and their contemporaries' had large ramifications in not only mathematics, but also in philosophy and semantics. Indeed the year after Hilbert published his foundation of classical mathematics, the philosopher Wittgenstein wrote extensively on the limitations of language, making the famous comment:

> *The limits of my language, mean*
> *the limits of my world.*

If mathematics is our language, where do you think we are limited?

References

1. P. Davis, R. Hersh, *The Mathematical Experience*, Pelican Books, New York (1981)
2. M. Giaquinto, *The Search for Certainty – A Philosophical Account of Foundations of Mathematics*, OUP (2006)
3. J. Hintikka, *The Philosophy of Maths. Languages in which Self Reference is possible*, R. M. Smullyan, OUP (1969)

2 0

Alan Turing Year
All 2012

100 years since this great British mathematician and computer scientist was born. Known for his contributions to computability theory, cracking the as Enigma in World War II and the Turing machine, he died at the age of 41.

Shaw + Physics Prizes
May and July 2012

Russian Fields Medallist Maxim Kontsevich received the 2012 Shaw Prize in Mathematical Sciences. He was also one of the 9 recipients of the Fundamental Physics Prize awarded for the first time this year.

Crafoord Prize
January 2012

Awarded to the Fields Medallists Terence Tao and Jean Bourgain.

Year of Mathematics
All 2012

…in India as a tribute to Srinivasa Ramanujan, as well as in Nigeria.

The Abel Prize
March 2012

Awarded, by the King of Norway, to the Hungarian-American mathematician Endre Szmeredi *"for his fundamental contributions to discrete mathematics and theoretical computer science, and in recognition of the profound and lasting impact of these contributions on additive number theory and ergodic theory."*

Smallest number with 10 divisors. Highly composite number. Double factorial of 6.

1 2

Discovery of the Higgs Boson

4 July 2012

CERN announced findings of a new particle that was consistent with the predicted Higgs boson. Evidence for this so-called 'God Particle' has been sought after for many decades, and if this really was a Higgs, its existence would explain many mysteries of the universe, including how matter attains mass.

21 / 12 / 2012

According to widely accepted arithmetic and astronomical theories, the world will … possibly … probably … not end.

ABC Conjecture

August 2012

Japanese mathematician Shinichi Mochizuki has published a proof online. It is still under verification.

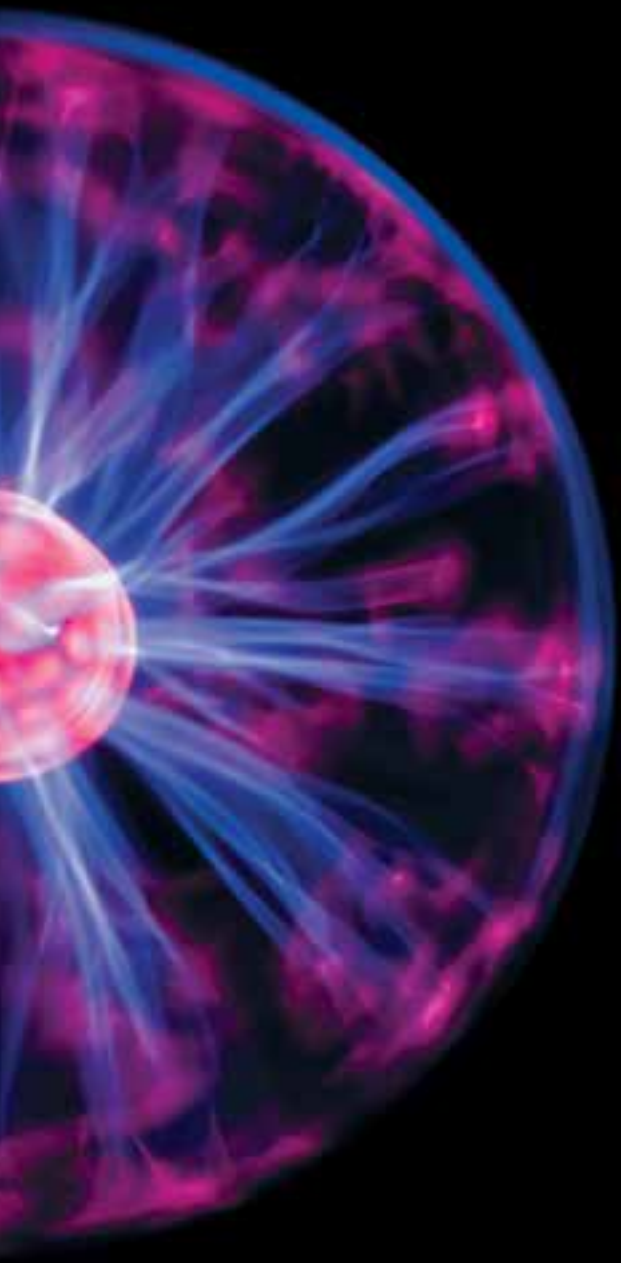
Let a, b, c be relatively prime integers with $a + b = c$. *Then for any* $\varepsilon > 0$, *there is some* C_ε *such that* $\max(|a|, |b|, |c|) \leq C_\varepsilon \prod_{p|abc} p^{1+\varepsilon}$ *when p is prime.*

Far reaching consequences include Roth's Theorem, Fermat's Last Theorem and the Mordell conjecture.

Junk DNA

September 2012

OK – not mathematics but a great discovery. While over 98 % of the human genome had been previously thought to serve no purpose, and so called junk DNA, the ENCODE project released 30 papers disproving this to reveal that over 80% perform vital functions in the body.



Hopes and Fears

Paul Dirac

First published in issue 32, 1969

A research worker who is actively following up some idea referring to the fundamental problems of physics has, of course, great hopes that his idea will lead to an important discovery. But he also has great fears – fears that something will turn up that will knock his idea on the head and set him back to the starting point in his search for a direction of advance. Hopes are always accompanied by fears, and in scientific research the fears are liable to become dominant.

As a result of these emotions the research worker does not proceed with the detached and logical mind that one would expect from someone with scientific training, but is subject to various restraints and inhibitions which obstruct his path to success. He may delay taking some step liable to force a rapid show-down, and may prefer first to nibble at side-issues that provide a chance of achieving some minor successes and gaining a little strength before facing the crisis.

For these reasons the innovator of a new idea is not always the best person to develop it. Some other person without the fears of the innovator can apply bolder methods and may make a more rapid advance. In the following there will be some examples that illustrate this situation.

Anyone who has studies special relativity must have wondered why it was that Lorentz, after he had obtained correctly all the equations of the Lorentz transformation, did not then take the perfectly natural step of considering all frames of reference to be on the same footing and so ar-

riving at the relativity of space and time. History does not record just what it was that held Lorentz back, but it can only have been some kind of fear, perhaps a subconscious one. He did not dare to venture out into a domain of thought completely foreign to anything that anyone had ever imagined. He preferred to remain on the solid round of mathematical transformations, where his position was unassailable. It needed the boldness of a younger man such as Einstein to take the plunge into a new domain.

The innovator of our present quantum mechanics was Heisenberg. At a time when atomic physicists were floundering about with the orbits of Bohr-Sommerfeld theory and feeling the need for a drastic alteration of basic principles, Heisenberg has the brilliant idea of constructing a new theory entirely in terms of observable quantities, quantities connected with observations on spectra. These are each connected with two atomic states, and the natural way of expressing them is in the form of matrices. Thus Heisenberg was led to consider matrices as dynamic variables.

He had not proceeded far in developing this idea before he noticed that his dynamical variables would not satisfy the commutative law of multiplication. This was most disturbing. It was inconceivable to a physicist in those days that dynamical variables could be any other than ordinary algebraic quantities, and with the appearance of non-commutation Heisenberg had grave fears that his whole beautiful idea would have to be given up.

When I read Heisenberg's first paper on the subject, I had the advantage over him in not having his fears, as it was not my own idea that was at stake. I was therefore able to look at the question from a more detached point of view.

I needed only a week or two to realize that the non-commutation which alarmed Heisenberg was really the dominating feature of the new theory. The idea of building up a theory entirely in terms of experimentally observed quantities, although a very pleasing philosophical doctrine, was of only secondary importance for the purpose of establishing a new dynamics.

My early work on quantum mechanics was thus concentrated on the problem of bringing non-commutation into dynamical theory. It was not really very difficult, because the previous atomic theory, the orbit theory of Bohr and Sommerfeld, was based on a form of dynamics, Hamilton's form, which turned out to be specially suitable for adapting to non-commutative algebra.

Heisenberg continued to develop his theory, in collaboration with other people in Göttingen. I worked independently from them, apart from getting the initial idea from Heisenberg. We published papers at about the same time, setting the foundations for quantum mechanics. Our styles were different on account of the different points of view we held, mine being based on non-commutation and Heisenberg's on the use of matrices built up from observable quantities.

Quantum mechanics was discovered quite independently by Schrödinger, working on entirely different lines. He had his own difficulties. He was thinking over the mathematical connection between waves and particles that had been discovered some time previously by de Broglie, and eventually found a way of generalizing it to apply to an electron moving in an electromagnetic field. He then had a very beautiful wave equation, conforming to relativity. He proceeded to apply it to the hydrogen atom and his worst fears were realized. The results did not agree with observation.

We know now that the discrepancy was due to the spin of the electron, which was unknown to Schrödinger at the time, although the experimentalists had begun to suspect it. It was a most depressing situation for Schrödinger, and led him to abandon the work for some months, and eventually to publish it only in the non-relativistic approximation, in which the discrepancy does not show up. The relativistic equation was later rediscovered by Klein and Gordon, who were not afraid to publish an equation in disagreement with observation, while Schrödinger was. So the equation now bears their name. It has some value in describing spinless mesons.

Schrödinger's quantum mechanics was soon found to be equivalent to that originated by Heisenberg, in spite of their first seeming so different. The basic equations of the new mechanics were securely established, and it became necessary to find a physical interpretation for them. With non-commutative algebra it could not be as direct as in the classical theory. The general physical interpretation was found to be only a statistical one. One could calculate probabilities, but could not usually predict an event with certainty.

A difficulty now appeared in connection with the relativistic equation of Klein and Gordon. The theory sometimes gave negative probabilities. It was a satisfactory theory only when it was used non-relativistically. I puzzled over this for some time and eventually thought of a new wave equation which avoided the negative probabilities. I found that it also gave automatically the spin of the electron, a most gratifying result. I proceeded to apply the new equation to the hydrogen atom, taking into account the relativistic corrections only to the first order of accuracy to simplify the calculations, and found agreement with observation.

The natural thing to do at this stage would have been to continue to higher orders of accuracy, but I did not do so. I was scared that they might not agree with observation. I hastily wrote up a paper with merely the first order of accuracy and published that. In doing so I felt I was consolidating a limited success, and even if the higher orders did go wrong there would still be something to stand on. It was left to Darwin, who did not share my fears to carry out the calculation to all orders of accuracy and see that the results were alright.

In my first paper on the subject (Proc. Roy. Soc A 117, page 610) there occurs the equation

$$F = \left(\frac{W}{c} + \frac{e}{c}A_0\right)^2 + \left(p + \frac{e}{c}A\right)^2 + m^2c^2.$$

The relativist, if he sees this equation nowadays, will say at once: there is a mistake here. The plus signs before the second and third terms on the right should be minus's. He will wonder how such a conspicuous mistake could have remained undetected in the proof-reading. He will wonder still more when he sees the same mistake perpetuated in later equations.

The explanation is that there is really no mistake and things were published as the author intended. The plus signs were the expression of a fear. At that time relativity was still unfamiliar and people had continually to cling to the symmetry of space and time so as not to let it out of their heads. The symmetry becomes perfect only if one uses a time variable which is $\sqrt{-1}$ times the usual time and makes a corresponding change in all 4-vectors. With this notation there are no mistakes in the paper. This notation was frequently used in those days, and it was not considered necessary to explain it every time it was used, because the context made it clear. The arrival of the new wave equation rather forced one to give it up, as it then became too clumsy.

The new wave equation led to a difficulty in that it allowed states of negative energy for the electron. Negative energies are never observed, but they could not be ignored in the theory. I thought of a way of coping with them, namely, to assume that in the physical world all or nearly all of the negative-energy states are occupied, so that ordinary positive-energy electrons cannot jump into them. An unoccupied negative-energy state is a hole which appears as a particle with a positive energy and a positive charge.

Right from the beginning I had the feeling that there would be symmetry between the holes and the electrons. This feeling was strengthened by the knowledge that in the chemical theory of the valency of atoms, there is a considerable amount of symmetry between an electron lying outside the closed shells and a hole in a closed shell. I did not want the symmetry. At that time it was believed that all positive charges were in protons, and the proton was much heavier than the electron. So I struggled with the hope that in some way the Coulomb interaction between the electrons would lead to a dissymmetry between the holes and the electrons, and was afraid that if this hope should fail the whole idea would have to be abandoned. It was left to others, in particular Weyl and Oppenheimer, to make the bold assertion that mathematical symmetry demanded that the holes should have the same mass as the electrons.

With these developments the theory of single particles was put into order. There remained problems concerned with interaction. If one sets up precise relativistic equations one finds that the interaction is so violent that the equations do not have any solutions. The difficulties are still not satisfactorily resolved and point to the need for some further drastic change in the foundations of atomic theory.

APPLY YOURSELF

Metaswitch is a global company that is empowering a new generation of communications applications and infrastructure. World-class products, services and networks are 'Built on Metaswitch': Your career can be too. We are growing strongly in an exciting market. Work with us and you'll see the effects of your work in how the whole world communicates. There's never been a better time to join!

Metaswitch Networks

careers.metaswitch.com

THE SUNDAY TIMES 100 BEST COMPANIES TO WORK FOR 2012

The Whirlpool Galaxy M 51a, as seen by Hubble

54 Number of coloured squares on a Rubik's cube. Perfect round on a par 72 golf course.

Quantum Gravity

Stephen Hawking, DAMTP Cambridge

First published in issue 32, 1969

The interactions that one observes in the physical universe are normally divided into four categories according to their botanical characteristics. In order of strength they are, the strong nuclear forces, electromagnetism, the weak nuclear forces and, the weakest by far, gravity. The strong and weak forces act only over distances of the order of 10^{-13} cm or less and so they were not discovered until this Century when people started to probe the structure of the nucleus. On the other hand electromagnetism and gravity are long range forces and can be readily observed. They can be formulated as classical, i.e. non quantum, theories. Gravity was first with the Newtonian theory followed by Maxwell's equations for electromagnetism in the 19th Century. However the two theories turned to be incompatible because Newtonian gravity was invariant under the Galilean group of transformations of inertial frames whereas Maxwell equations were invariant under the Lorentz group. The famous experiment of Michelson and Morley, which failed to detect any motion of the Earth through the luminiferous aether that would have been required to maintain Galilean invariance, showed that physics was indeed invariant under the Lorentz group, at least, locally. It was therefore necessary to formulate a theory of gravity which had such an invariance. This was achieved by Einstein in 1915 with the General Theory of Relativity.

General Relativity has been very successful both in terms of accurate verification in the solar system and in predicting new phenomena such as black holes and the microwave background radiation. However, like classical electrodynamics, it has predicted its own downfall. The trouble arises because gravity is always attractive and because it is universal i.e. it affects everything including light. One can therefore have a situation in which there is such a concentration of matter or energy in a certain region of space-time that the gravitational field is so strong that light cannot escape but is dragged back. According to relativity, nothing can travel faster than light, so if light is dragged back, all the matter must be confined to a region which is steadily shrinking with time. After a finite time a singularity of infinite density will occur.

General Relativity predicts that there should be a singularity in the past about 10,000 million years ago. This is taken to be the "Big-Bang", the beginning of the expansion of the Universe. The theory also predicts singularities in the gravitational collapse of stars and galactic nuclei to form black holes. At a singularity General Relativity would lose its predictive power: there are no equations to govern what goes into or comes out of a singularity. However when a theory predicts that a physical quantity should become infinite, it is generally an indication that the theory has broken down and has ceased to provide an accurate description of nature. A similar problem arose at the beginning of the Century with the model of the atom as a number of negatively charged electrons orbiting around a positively charged nucleus. According to classical electrodynamics, the electrons would emit electromagnetic radiation and would lose energy and spiral into the nucleus, producing a collapse of the atom. The difficulty was overcome

by treating the electromagnetic field and the motion of the electron quantum mechanically. One might therefore hope that quantisation of the gravitational field would resolve the problem of gravitational collapse. Such a quantisation seems necessary anyway for consistency because all other physical fields appear to be quantised.

So far we have had only partial success in this endeavour but there are some interesting results. One of these concerns black holes. According to the Classical Theory the singularity that is predicted in the gravitational collapse will occur in a region of space-time, called a black hole, from which no light or anything else can escape to the outside world. The boundary of a black hole is called the event horizon and acts as a sort of one way membrane, letting things fall into the black hole but preventing anything from escaping. However, when quantum mechanics is taken into the account, it turns out that radiation can "tunnel" through the event horizon and escape to infinity at a steady rate. The emitted radiation has a thermal spectrum with a temperature inversely proportional to the mass of the black hole. As the black hole emits radiation, it will loss mass. This will make it get hotter and emit more rapidly. Eventually it seems likely that the black hole will disappear completely in a tremendous final explosion. However the time scale for this to happen is much longer than the present age of the Universe, at least for black holes of stellar mass, though there might also be a population of much smaller primordial black holes which might have been formed by the collapse of irregularities in the early Universe.

One might expect that vacuum fluctuations of the gravitational field would cause "virtual" black holes to appear and disappear. Particles, such as baryons, might fall into these holes and be radiated as other species of particles. This would give the proton a finite lifetime. However it is difficult to discuss such processes because the standard perturbation techniques, which have been successful in quantum electrodynamics and Yang-Mills theory do not work for gravity. In the former theories one expands the amplitudes in a power series in the coupling constant. The terms in the power series are represented by Feynmann diagrams. In general these diverge but in these theories all the infinities can be absorbed in a redefinition or "renormalisation" of a finite number of parameters such as coupling constants as masses. However in the case of gravity, the infinities of different diagrams are different and so they would require an infinite number of renormalisation parameters whose values could not be predicted by the theory. In fact the situation is not really that much worse than with the so-called renormalisable theories since even with them the perturbation series is only asymptotic and does not converge, leaving the possibility of adding an arbitrary number of exponentially vanishing terms with undetermined coefficients.

The problem seems to arise from an uncritical application of perturbation theory. In classical general relativity it has been found that perturbation expansions around solutions of the field equations have only a very limited range of validity. One cannot represent a black hole as a perturbation of flat space-time yet this is what summing Feynmann diagrams attempts to do. What one needs is some approximation technique that will take into account the fact that the gravitational field and the space-time manifold can have many different structures and topologies. Such a technique has not yet been developed but we, at Cambridge, have been approaching the problem by studying the path integral approach formula-

Sum of the first six triangular numbers and the first six odd primes. Tetranacci and Pronic number.

tion of quantum gravity. In this the amplitudes are represented by an integral over all metrics

$$\int D[g] \exp(-\hat{I}[g])$$

where $D[g]$ is some measure on the space of all metrics g and $\hat{I}[g]$ is the action of the metric g.

If the integral is taken over real physical metrics (that is, metrics of Lorentzian signature $- + + +$), the action I is real so the integral oscillates and does not converge. To improve the eigenvalues one does a rotation of 90° in the complex t-plane. This makes the metric positive definite (signature $+ + + +$ and the action I pure imaginary so that the integral is of the form

$$\int D[g] \exp(-I[g])$$

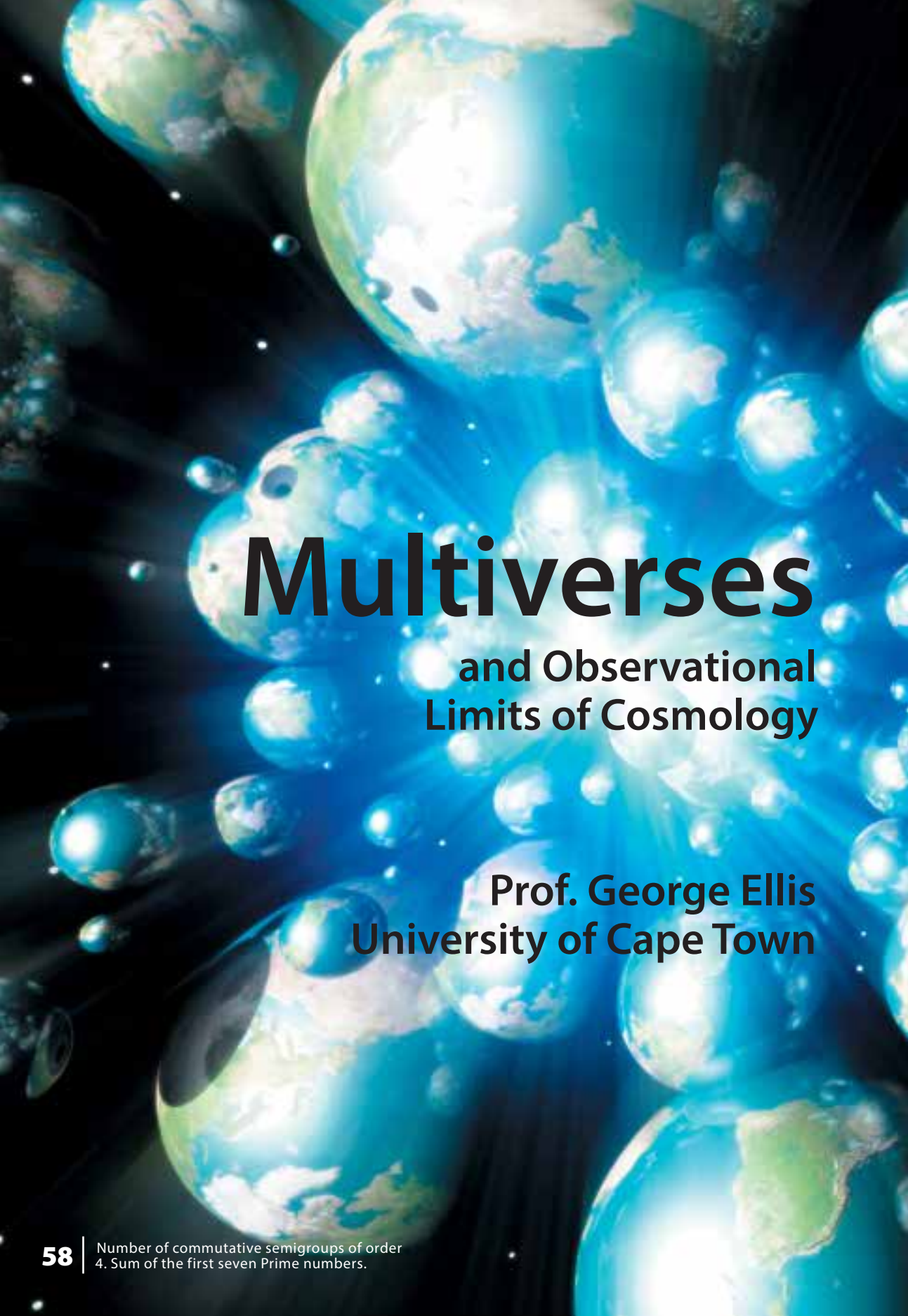
where $\hat{I} = -iI$. The Euclidean action $\hat{I}$ has certain positive definite properties.

One is thus led to the study of positive definite metrics (particularly solutions of the Einstein equations) on four-dimensional manifolds. If the manifolds are simply connected, their topology can be classified (at least up to homotopy) by two invariants, the Euler number as measuring the number of *holes* or *gravitational instantons* and the signature measures the difference between right-handed instantons and left-handed ones. It seems that the dominant contribution to the path integral comes from metrics with about one instanton per Planck volume 10^{-142} cm^{3}s. Thus space-time seems to be very highly curved and complicated on the scale of the Planck length 10^{-33} cm, even though it seems nearly flat on larger scales.

However we still do not have a proper scheme for evaluating the path integral. The difficulty lies in defining a measure $D[g]$ on the space of all metrics. In order to obtain a finite answer it seems necessary to make infinite subtractions and these leave finite undetermined remainders. There is a possible way of overcoming this difficulty which may come from an extension of General Relativity called supergravity. In this the spin 2 graviton is related to a spin 3/2 field and possibly fields of lower spin by anticommuting "supersymmetry" transformations. In these theories there is an equal number of bosons (integer spin particles) and fermions (half integer spin particles). The infinities that arise in the path integral from the integration over boson fields seem to cancel when the infinities that arise from the integration over the fermion fields, raising the hope that one could provide a proper mathematical definition of the path integral, maybe some limiting process.

Supergravity theories have another very desirable feature, they may unify gravity with the other interactions and particles in physics. In 1967 Salem and Weinberg proposed a unified theory of the electromagnetic and weak interactions. This has had considerable success in predicting experimental results though the final confirmation will have to wait for the next generation of particle accelerators. Nevertheless, it has given great stimulus to attempts to unify the strong, the weak and the electromagnetic interactions into a "Grand Unified Theory". A feature of such theories is that the complete unification is seen only at the very high energies of the order of 10^{19} Gev, at which quantum gravitational effects should become important. It may well be therefore that one will be able to achieve the unification only by incorporating gravity as well in a completely unified theory which would describe all of physics. This was the goal to which Einstein devoted the last thirty years of his life, without much success. The prospects look brighter now though it is still probably quite a long way off.

The 57-cell is a self-dual abstract regular four-dimensional polytope. It has symmetry group $L_2(19)$.



Multiverses

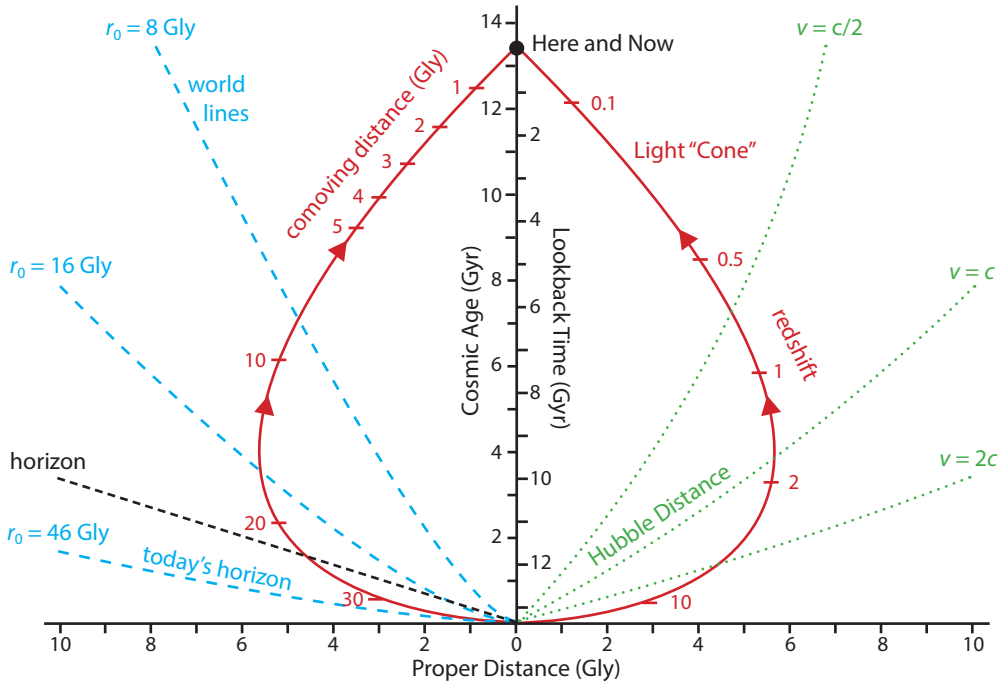
and Observational Limits of Cosmology

Prof. George Ellis
University of Cape Town

There has been a recent flurry of articles and books proposing that we live in a multiverse rather than a universe: there is not one universe but many (Rees 1999, Rees 2000, Carr 2008, Greene 2011). This raises key issues about the validity and utility of mathematical models, and their relation to what exists. Mathematical models must be coherent as models, based in whatever the underlying physics is; but that is not by itself sufficient to make them physically relevant. If they are meaningful as mathematical models of some physical system, they must be applicable to that context; which means you need to be able to test them and see if they describe the system well. If there is no possible way to test them, you have a problem: it is unclear whether they are indeed reliable models of reality. And that is a major issue that arises as regards multiverse theories .

A variety of kinds of multiverses have been envisaged by many authors. In his recent book The Hidden Reality (Greene 2011), Brian Greene proposes nine different types of multiverse theories:

1. Existence beyond the horizon: Invisible parts of our own universe.
2. Chaotic inflation, leading to different expanding domains in separate places.
3. Brane worlds of M-theory (Four-dimensional space-times embedded in higher dimensional spacetimes).
4. Cyclic universes, leading to different expanding domains at different times.
5. The Landscape of string theory embedded in a chaotic cosmology.



▲ **Figure 1** Space Time Diagram – Normal Distance and Time

6. The Everett quantum multi-universe: other branches of the wave function.
7. Holographic projections (currently a trendy proposal in cosmology).
8. The universe is a computer simulation.
9. All that can exist must exist—the "grandest of all multiverses", the separate universes being totally disjoint from each other.

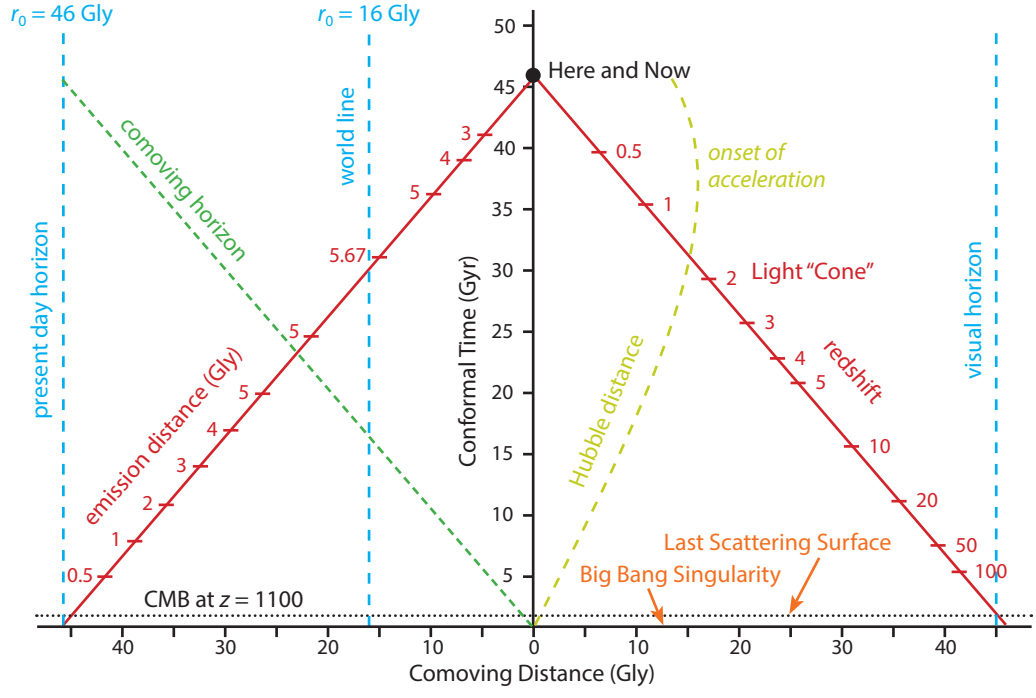
Now one thing is clear – they can't all be true, for they conflict with each other. There remains the final possibility:

10. Maybe none of them is true – there is just one universe.

I will concentrate on the most popular one: chaotic inflation (2), usually coupled with the landscape of string theory (5). I will show firstly that there is no way to directly verify that this model is true, and secondly that it is not based in well understood and verified physics. Hence while it may possibly be true, it has not been proved so, and indeed that proof may well be impossible. The reason we can't prove a multiverse exists observationally is due to the nature of its spacetime structure, which on a large scale is governed by Einstein's General Relativity Theory. When we model the large scale structure of the universe, our cosmological models are surprisingly simple: they assume a basic structure that is both spatially homogeneous (all physical quantities are the same everywhere at the same cosmic time) and spatially isotropic (there are no preferred directions in the sky when we average matter on large enough scales). This geometry is represented by the metric of the spacetime (see Appendix), which has a scale factor $a(t)$ representing the change of distance between galaxies with time, whose time evolution is determined by the Einstein's gravitational field equations, depending on the matter and radiation content of the universe. The metric also determines the paths of photons through spacetime, and so in particular determines the size of the visual horizon as a function of cosmic time.

To understand this properly one of course needs to contemplate the equations of the theory, given in the Appendix. However we can also understand its relevant properties straightforwardly from spacetime diagrams showing how causal relations work in these models. The way such diagrams work, and their relation to the under-

Order of the smallest non-abelian simple group.
Number of minutes in one hour and one degree.



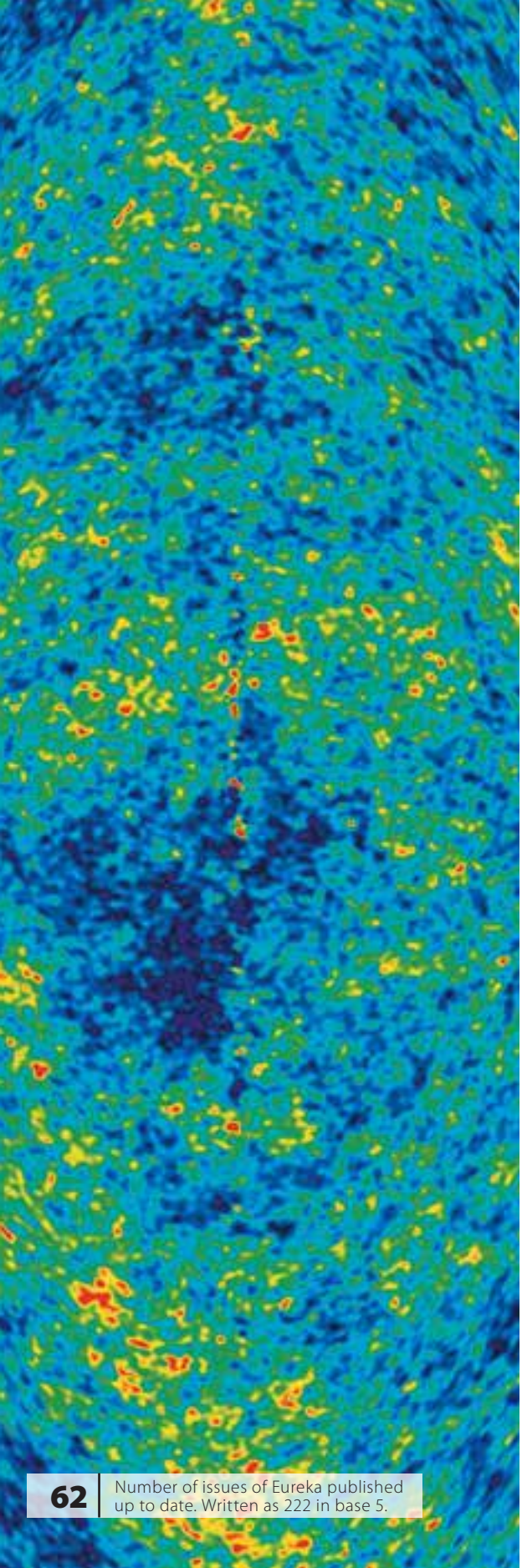
▲ **Figure 2** Space Time Diagram – Comoving Distance and Conformal Time

lying spacetime metric, is discussed in detail in my book with Ruth Williams (Ellis and Williams 1995). The relevant details are as follows.

Figure 1 is a space time diagram representing spatial distances and cosmic time correctly. Time is plotted vertically, and distance horizontally. The start of the universe is at $t = 0$; galaxy world lines diverge from each other since then. Our galaxy world lie is at the centre ($r = 0$); the present is labelled "here and now". Our past light cone is marked in red; this is the path through spacetime of photons that are reaching us now. Going back into the past, it reaches a maximum radius and then contracts back to the big bang singularity; this is basically because gravity bends light – one of Albert Einstein's major discoveries.

Now the problem with that diagram is we can't see causal relations near the big bang very well. We can correct that by changing to stretched distance and time coordinates, that transform the past light cones to lines at ± 45° and matter world lines to vertical lines (Figure 2). This is allowed because Einstein's theory allows the use of any coordinates whatever to represent a given spacetime (this is the principle of general covariance). The initial singularity – a point in Figure 1 – is then represented by the horizontal line at the bottom. One should note that this singularity is not part of spacetime – it is the boundary of spacetime. This diagram also shows something not represented in Figure 1: the dotted horizontal line just above the singularity. This represents the surface of last scattering ("LSS"), where matter and radiation decoupled from each other in the early universe. This is the furthest back that we can see, because the universe was opaque to radiation at earlier times. Hence any earlier physics – the way the universe was created, the subsequent inflationary era - is not visible to us (this is similar to the way the surface of the Sun hides its interior from us). Most importantly, whatever that earlier physics was does not affect light propagation since decoupling of matter and radiation at the LSS: hence the causal limits on what we can see are unaffected by any such earlier physics.

Now the key point is that there is a furthest set of matter we can see by any electromagnetic radiation whatever; its world lines are marked as the Visual Horizon on the right. It is the world lines of matter that pass through the intersection of our past light cone with the LSS. This matter emit-



ted the cosmic blackbody background radiation detected by the WMAP satellite, so the famous microwave background anisotropy map (Figure 3) is just the image we get of this matter at the LSS. Any further out matter cannot be seen or detected by us by means of any radiation whatever (assuming no radiation moves faster than light – a key feature of relativity theory). The causal horizon (marked here as "present day horizon") lies further out, and depends on early physics.

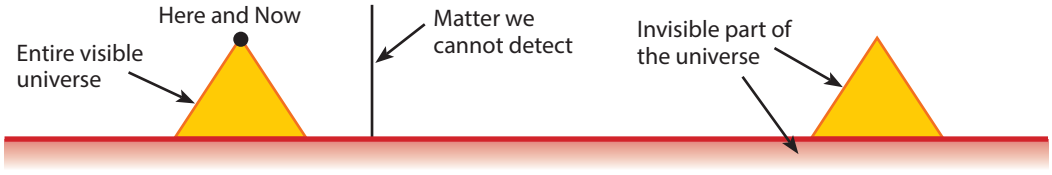
In order to get a better picture of our observational limits, we need to step back a bit and see the bigger context for Figure 2, depicted in Figure 4: the whole of Figure 2 is the left hand triangle there. We cannot detect matter outside there by any means whatever. Hence we have no means of telling what conditions are like inside the presumed universe domain on the right (the same size as our entire visible universe domain on the left). Physics there might be the same as here, or it might be totally different. There is simply no way we can ever find out.

So here is the basic problem for multiverse proponents: no observational data whatever are available to verify their claims of distant universe domains out there with different physics than in our domain. If the basis of science is verifying theories by observation, then multiverse theories are not science. The assumption made in those theories is that we that can extrapolate to 100 Hubble radii, 101 000 Hubble radii, or much much more (the word 'infinity' is casually bandied about) to determine in broad terms what conditions are like there. That's not testable science.

But there is another line of argument. Maybe one can justify the multiverse assumption if is a necessary outcome of known and tested physics, even if one cannot directly verify its existence. This is indeed a sound line of reasoning. The problem with it is that several aspects of the physics supposed to lie the multiverse are hypothetical rather than well established: they are major extrapolations of known physics into the unknown, and those extrapolations may or may not be true. This issue is discussed in depth by Banks (2012), who shows quite clearly that none of the supposed physics (Coleman-de Luccia tunneling, the landscape of string theory, the supposed connection between chaotic inflation and string theory vacua) is well established physics.

A third line of argument is that existence of a multiverse explains anthropic coincidences in

◀ **Figure 3** The Cosmic Background Radiation sky – our image of the Last Scattering Surface.



▲ **Figure 4** The entire visible universe is a tiny fraction of the claimed multiverse. Most of its regions (if they exist) are not observationally accessible to us by any means.

cosmology: why the universe is a suitable place for life to exist, in particular explaining the value of the cosmological constant (the "dark energy" currently causing the universe to accelerate). This case is made for example by Martin Rees (1999, 2001), see also Carr (2009). Now it does indeed provide such an explanation. Does this therefore justify belief in a multiverse? Yes if you think theory trumps observational testing in a scientific theory; no otherwise. Key philosophical issues about the nature of scientific theories underlie this choice; a discussion is given in Ellis (2006).

There are however two exceptions to this gloomy picture re testability of the multiverse idea. The first is the possible existence of "small universes": universes where the spatial sections are spatially closed on a scale smaller than that of the visual horizon (Lachieze-Ray and Luminet 1995). In that case the horizontal axis of Figure 2 would close on itself on a scale smaller than that of the visual horizon, and we would already have seen all the matter there is in the universe, thus disproving the multiverse hypothesis. This intriguing possibility can be tested in various ways, in particular by searching for identical circles of temperature fluctuations in the CMB sky. This search has so far proved unsuccessful: this remains a possibility, but is perhaps unlikely.

The second exception would be if there were collisions between different bubbles in the multiverse, resulting in detectable disk-like patterns in the CB sky. If such bubble collisions were detected and additionally could be associated with a variation of physics in the different bubbles, for example different values of the fine structure constant, this could legitimately be taken as vindication of the physics supposed to underlie the multiverse proposal. This has so far not been observed.

A final comment relates to the issue of infinities. It is often said that infinities of universes occur in the multiverse (see for example Vilenkin 2007). This is a very dubious claim. Firstly, David Hilbert has stated "the infinite is nowhere to be found in reality, no matter what experiences, observations, and knowledge are appealed to." (Hilbert 1964). I strongly concur. Secondly, in any case such claims are not verifiable, for there is no possibility whatever of verifying them (no matter how many entities you have counted, you have not proved an infinity exists). If science is to do with testable claims, then any such claims are not science.

For other motivations for the multiverse, arguments in its favor, and counterarguments, see my Scientific American article (Ellis 2011), the book edited by Carr (2009), and Kragh (2012).

References

1. T Banks, *The Top 10500 Reasons Not to Believe in the Landscape*, arXiv: 1208.5715 (2012)
2. B Carr, *Universe or multiverse?*, CUP (2009)
3. G F R Ellis,*Issue in the Philosophy of cosmology*, Handb. in Philosophy of Physics (2006)
4. Butterfield, Earman, arXiv: 1183.1285 (2006)
5. G F R Ellis, *Does the multiverse really exist?* Scientific American 305, 38-43 (2011)
6. G F R Ellis, T. Rothman, *Lost Horizons*, American Journal of Physics 61-10 (1993)
7. G F R Ellis, R M Williams, *Flat and Curved Spacetimes*, Oxford University Press (2000)
8. B Greene, *The hidden reality: Parallel universes and the deep laws of the cosmos*, Knopff (2011)
9. D Hilbert, *On the infinite*, in *Philosophy of mathematics*, Prentice Hall (1964)
10. H Kragh, *Criteria of Science, Cosmology, and Lessons of History*, arXiv: 1208.5215v1 (2012)
11. M Lachieze-Ray & J P Luminet, *Cosmic topology*, Physics Reports 254, 135 (1995)
12. M Rees, *Just six numbers*, Weidenfeld and Nicholson (1999)
13. M Rees, *Our cosmic habitat*, Princeton University Press (2001)
14. A Vilenkin, *Many worlds in one: The search for other universes*, Hill and Wang (2007)

64 Smallest number with 7 divisors. Index of Graham's number in the sequence 3, 27, 7625597484987, …

Primes and Particles

Jack Williams, Clare

The long-studied and incredibly elusive Riemann zeta-function has baffled number theorists for centuries. Deeply connected to the prime numbers and their distribution among the integers, its importance in number theory is well known. The Riemann hypothesis is perhaps the most famous unsolved problem in mathematics, not least because the Clay Mathematics Institute has offered $1 million for a solution.

However, interest in this mysterious function extends far beyond esoteric results in number theory. With implications for physics, probability and statistics, it is more than just a number theoretic curiosity. The underlying distribution of the zeros along the critical line $\Re(s) = \frac{1}{2}$ penetrates many branches of mathematics and there has been growing interest in the obscure connections it reveals between these fields.

When defined as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

for $\Re(s) > 1$, it seems unconnected to the primes. But Euler first noticed the factorisation

$$\zeta(s) = \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \cdots\right)\left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \cdots\right)\left(1 + \frac{1}{5^s} + \frac{1}{5^{2s}} + \cdots\right)\cdots$$

which holds by the fundamental theorem of arithmetic. Summing the geometric series yields

$$\zeta(s) = \prod_{p\ prime} \frac{1}{1 - p^{-s}}$$

This factorisation, which can be made rigorous, reveals a deep connection with the primes.

As an illustration of the zeta-function's role in number theory, we give simple proof of the familiar Euclidean theorem.

Theorem There are infinitely many primes.

Proof: Suppose not. Then taking $s \to 1^+$ in the identity,

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \equiv \prod_{p\ prime} \frac{1}{1 - p^{-s}}$$

gives

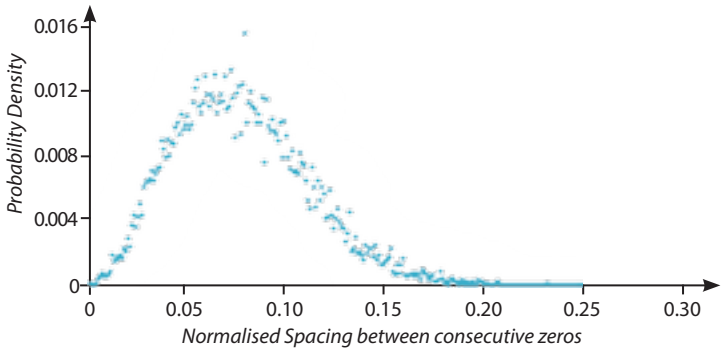
$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p\ prime} \frac{1}{1 - \frac{1}{p}}$$

which must be finite because there are only finitely many terms in the right hand side. But since the harmonic series diverges, this is a contradiction. □

Scattered recklessly throughout the integers, the primes have been studied for thousands of years. Gauss' prime number theorem states that

$$\pi(n) \sim \frac{n}{\ln(n)}$$

where $\pi(n)$ is the number of primes less than or equal to n. This means that the probability that is prime is roughly $\frac{n}{\ln(n)}$ for large n. Here n need not be very large for a good approximation to $\pi(n)$. For $n = 100$, it estimates about 21 primes when the correct value is 25.



◀ Probability density function for the normalised spacing between consecutive zeroes of the Riemann zeta-function

In 1859, Riemann published another link between the zeta-function and the primes, giving an exact formula for $\pi(n)$, which improved on the prime number theorem. Riemann's formula uses the location of zeroes in the analytic continuation of the Riemann zeta-function. Number theorists are interested in it because the distribution of zeroes gives information about the distribution of primes.

While prime numbers are sparser among the larger integers, the Riemann zeroes become denser further from the real line. It is possible to show that

$$n(T) \sim \frac{T \ln(T)}{2\pi}$$

where $n(T) = \left|\left\{y \in \mathbb{R} : \zeta\left(\frac{1}{2} + yi\right) = 0, 0 < y \leq T\right\}\right|$. However, these zeroes are not arranged regularly along the critical line and the finer detail has been studied in more depth.

One way to get information about this distribution is to study a spacing distribution, the distribution of the distances between consecutive zeroes. These distances can be normalised to account for the 'global' distribution given by $n(T)$. The correct normalisation is $\hat{\gamma}_j = \frac{\gamma_j \ln(\gamma_j)}{2\pi}$, which results in the following distribution, estimated numerically using zeroes numbered $10^{21} + 1$ to $10^{21} + 10000$.

Surprisingly, this distribution arises naturally in other areas of mathematics and physics and seems to be fundamental to many seemingly unrelated problems. Experiments have been performed in which neutrons are scattered off a heavy nucleus. The resulting cross-section contains peaks, the scattering resonances, and troughs. If a neutron's energy is near to one of the peaks, it is repelled by the nucleus and if it is near a trough, it can pass through effortlessly. This strange behaviour leads to different scatterings. Although these problems are too difficult to solve analytically or numerically, empirical data can be obtained and a similar analysis to the separation of the zeroes can be performed on these peaks. Astonishingly, the probability density matches that of the Riemann zeroes. In fact, the density does not depend greatly on the particular nucleus being used. This remarkable connection is still poorly understood.

The mathematical structure which models the scattering resonances is also quite unexpected. In the 1950s, Eugene Wigner proposed a statistical model based on random matrices. Inspired by Heisenberg's formulation of quantum mechanics, in which properties of an atom or a nucleus can be described by a Hermitian matrix, he put forward the random-matrix conjecture. He suggested that these peaks follow the same spacing as eigenvalues of Hermitian matrices whose elements are chosen from some probability distribution, usually normal. The eigenvalues of the matrix are correspond to energy levels of the spectrum. Using such random matrices, one can obtain spacings that are statistically similar to both the neutron scattering data and the distribution of zeroes of the Riemann zeta-function.

Although it is possible that the zeroes play some direct role in physical systems, the unexpected connection between the primes, matrices and neutron scattering may reflect a deeper result, uniting many seemingly disjoint areas of mathematics. There has been progress in this direction. Theorems have been proved showing that there is a single limiting distribution for the eigenvalue spectrum arising from a large set of random matrices. Similar in spirit to the central limit theorem, these results mirror the type universality intrinsic to results which connect many areas of mathematics.

M-Theory, Duality and Art

Dr David Berman, Queen Mary University London

This will not be an article about mathematics or theoretical physics but more about why we carry out such activities and the common ground mathematicians share with those in the arts. When we look at the long history of mathematics there is a deep aesthetic sensibility present in the writings of mathematicians. Common references to beauty, symmetry emerge as almost a guiding tool. After all, mathematics is not just about what is true but what is interesting. What is interesting and beautiful seems somewhat subjective and yet when faced with the profound beauty of any number of mathematical theorems (the reader should think of her favourite theorem at this point) then it just seems so blindingly obvious that no explanation is ever needed. I am a theoretical physicist and I remember still the spine tingling excitement at first seeing Noether's theorem. This link between the symmetries of the laws of nature and conserved quantities is to me an example of the irresistible beauty that mathematics can uncover in nature.

A typical example of the aesthetic raptures of mathematicians is found in Poincare's famous quote, "The scientist does not study nature because it is useful. He studies it because he delights in it and delights in it because it is beautiful…I do not talk here of the type of beauty that strikes the senses but a profound beauty that comes from the harmonious order of the parts."

Poincare's "harmonious order of the parts" is somehow the mathematical beauty we know and love. His "beauty that strikes the senses" is I suppose a reference to more common ideas of beauty present in the visual arts. And yet are the two so necessarily different or can one reflect the other though perhaps with the inevitable loss of an imperfect reflection.

In a normal distribution, 68% of values are within one standard deviation of the mean.

When we look at some of the goals and ideas of artists there is often a theme uncovering different "ways of seeing" and of challenging our everyday view of the world. (This is a quick and perhaps over easy generalisation but let me continue). In the 20th century, the artist followed many of the revolutions of 20th century physics with a removal of direct representation and a strong foray into abstraction. In Mondrian's seminal essay, "Natural reality and abstract reality", the artist puts forward a manifesto justifying the abstraction of universals from nature, concentrating on essential building blocks and looking at structural connections while ignoring the detail present in the individual objects to bring out the common, shared structures in nature. His essay is comfortable reading for a contemporary theoretical physicist.

And so in goals and in a very human belief in aesthetics perhaps the differences are not so great after all. The language of mathematics is where barrier to communication lies. Its enormous power prevents the concepts from being accessible or visceral. The challenge then is to capture something of an idea or concept in physical work.

Duality is a key idea in mathematics and is now at the centre of the frontier in theoretical physics, string and M-theory. (One can argue whether duality really means the same thing in the context of mathematics and in theoretical physics but we will pass over this).

◀ "Generalised Geometry 1&2", Berman and Davey

▲ "125 GeV" in wood, Berman and Davey

In string theory, there is a duality symmetry known as T-duality. This is a fundamental ambiguity in the description of the space time background in which the string lives. If the space time has some specific properties (technically, it should possess an isometry and be compact so that its first homotopy class is nontrivial) then there will be two backgrounds that will be related to each other that in ordinary differential geometry will be inequivalent and yet will be indistinguishable from the point of view of the string. These pairs are known as T-duals. This duality is stringy in nature and leads on to the idea of stringy geometry that differs from our usual notion of geometry in that such ideas of T-duality get built in.

The Turner prize winning artist Grenville Davey, has spent many years working on sculptures that should be seen in isolation but as objects that bring out relationships and symmetries. They are collections of objects which when brought together show a relation.

Together Grenville and I have been exploring ways to have a sculptural manifestation of some of the ideas in theoretical physics such as T-duality and spontaneous symmetry breaking. The goal is not to explain or exemplify but simply to inspire.

Computer representations of T-dual manifolds though perhaps accurate are not what we were after. Instead the works were to be influenced by the mathematics rather than represent the mathematics in some faithful way. This allowed the process to be free and in the end driven by the detailed aesthetics of the pieces themselves than by faithfulness to an idea. The result has been a series of sculptures that have been shown in various gallery spaces but also in the Isaac Newton Institute for mathematical sciences.

▲ "125 Gev", Berman and Daveyy

▼ "Generalised Geometry", Berman and Davey

Who knows whether there is anything of T-duality or symmetry breaking in these works. What is interesting is the process by which the pieces came about and the fact that some very abstract mathematics has given rise to some pieces of sculpture and influenced the mind of a leading British sculptor.

tpp

systmone
ONEPATIENTONERECORD

Kick start a career in IT...

Join our team of Software Developers on a starting salary of £24K and excellent benefits

The requirements...Degree of 2:1 or higher, AAB at A-level, with an A in Maths

No experience required

To apply, send your CV and covering letter to careers@tpp-uk.com

Please visit our website for more information at www.tpp-uk.com

Contact 0113 20 500 82 **or email** careers@tpp-uk.com

Glacier Dynamics

Indranil Banik and Justas Dauparas

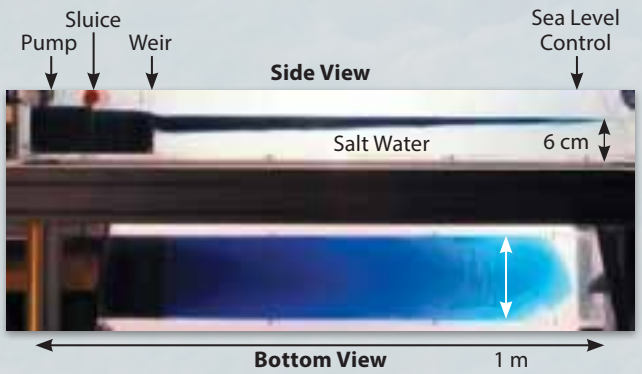
Climate change is now widely expected to cause significant changes to conditions on Earth in the next century, with our actions playing a key role in determining what happens next. One of the less well understood effects is sea level rise. This will likely be dominated by glacier retreat in Antarctica and Greenland.

It is these glaciers in particular that we attempted to understand this summer. Our method was to use a laboratory model for ice, which captures a hitherto neglected but we believe critical aspect: changes in viscosity. With a high rate of shear (velocity gradient), there is increased melting between adjacent crystals. This leads to a reduction in viscosity.

Our laboratory model for ice was Xanthan, a shear-thinning biological polymer. We started by considering the ice shelf, believing (or hoping for) the sheet to have been solved for already, as a viscous gravity current. The situation without sidewalls was simple enough for us to solve it analytically without experiments to guide us. We found that the front goes as a power law in time (it accelerates). The other key result was that, for a fixed source thickness and fixed entry flux, the thickness at any location does not change (once the front reaches it). However, the source thickness remained a mystery. The width is still unknown (though it increases if the flux is higher).

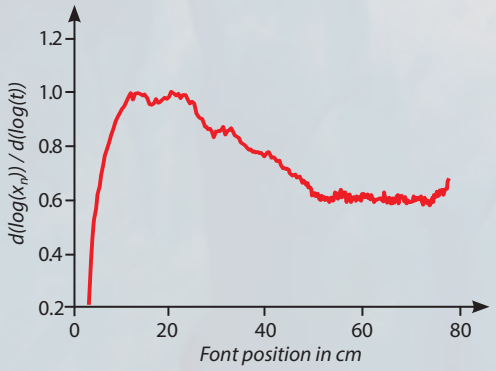
Happy with this very early (partial) breakthrough (the first real one in our careers), we then attempted to understand the effect of sidewalls. The motivation was ice shelves inside canyons, which is not at all unusual. Also common is slowly flowing pack ice completely filling a bay, leading to some friction at the edges. A typical experiment is shown in operating configuration, near the end of a run.

Almost everything was made inside the workshop of the GK Batchelor Laboratory. We pioneered a

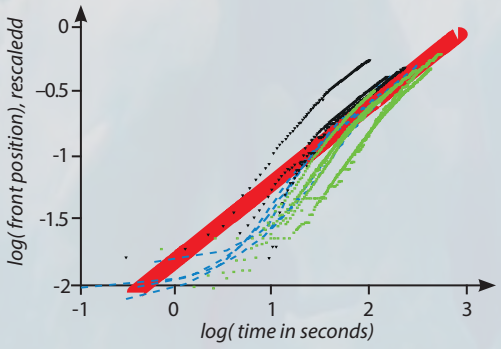


◀ **Figure 1** The basic experimental setup. The flow is from left to right – under the sluice, over the weir, into the ocean. This makes it more uniform across the channel – essential if we want good sidewall contact.

Average number of heartbeats per minute.
Percent of the human body consisting of water.



▲ **Figure 2** Shown here is the ratio of the fractional change in the front position to the fractional change in time, over a short period. Note that the tank is 15cm wide, so power-law behaviour starts at approximately three times this.



▲ **Figure 3** Front position plotted against time on logarithmic axes, with rescaling according to changes in flux and other parameters. The red band is our theory, allowing for errors. The blue and green data points are from experiments set up slightly differently, while the black ones are considered unreliable.

sea level control system, to stop large rises in sea level as Xanthan enters the ocean. Note the minimal drop in Xanthan from the weir into the ocean – this is because the sea level is < 1 mm below the top of the weir. Without removing saltwater from the ocean, we'd be forced to accept a 2 cm drop, potentially affecting the shelf over a large region. Also, conditions would alter significantly during the experiment (as the 2 cm gap goes down to 0). Fortunately, our control system was able to hold the sea level constant to within 1 mm.

Although there were other problems (like 3 specks of rust on the weir ruining 4 runs, until we realised); we mention the sea level in particular as we designed the system ourselves and because it wasn't controlled at all in a previous study, leading to it being fundamentally flawed.

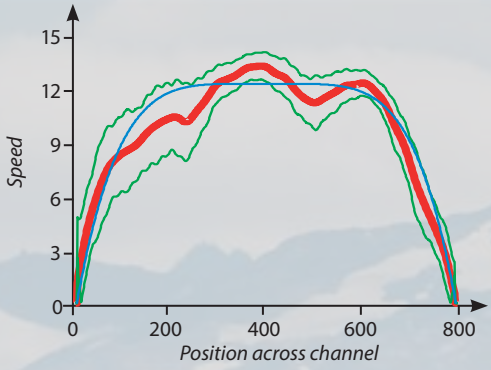
The key parameters (front position and peak thickness, near the source); are clearly visible in Figure 1. We took a photo like this every second. We speeded up the data analysis by using Matlab to trace the outlines (it's 2012, this sort of thing is easy). What we were looking for were clues – in particular, if the front was going as a power law in time. But there was another major problem, and our apparatus wasn't to blame.

The Xanthan was at a 1% concentration – this was too viscous. It didn't really have enough time to spread laterally, so the thickness at the walls was low, even in the best experiments. This meant that sidewall friction might not be dominating the system (compared to water pressure).

Although we got a tentative indication of power law-type behaviour, we weren't happy and reduced the concentration to 0.5% after a quick test. Then we had a real Eureka moment, though I didn't fully appreciate what it meant (unlike some of the people on the team). As can be seen below, the behaviour did indeed appear to converge to a power law. The uptick at the end is due to seawater extraction sucking the shelf with it (a little). As the front slows down, the flat region on the graph below actually lasts for 100 seconds (in a 250s experiment).

This was the defining moment of our project. We rapidly found more experiments indicating similar behaviour. Then, the quest was on to explain it – everyone was more confident than me that we would now be able to succeed for sure. The good thing was that the value of 0.6 above is only a bit different to what would have happened if xanthan was a Newtonian fluid (value = 0.67). However, I didn't share the enthusiasm that the shear-thinning nature of xanthan was having only a minor effect – maybe on the numbers, but I thought it made things fundamentally different.

We thought about it very carefully and eventually realised that, amazingly, the behaviour is indeed similar to the Newtonian case (where the flow is essentially a Poiseuille flow, with gradients in thickness driving it). Once the shelf is very long so sidewall friction dominates, then for our case, we get what we termed a generalised Poiseuille flow. The velocity is still polynomial in lateral position, but (for a shear-thinning fluid) there's a sharper edge (i.e. the power is more than 2).



▲ **Figure 4** Speed as a function of position across the channel. The theoretical curve is the smooth blue one. Raw data is in red, but errors are present so any value within the green curves is consistent with the data. Note that the tank is 15cm wide.

▲ **Figure 5** Side view of our shelf. Theory predicts the apparent discrepancy in the middle, where the gradient exceeds that near the source (right) – this should cause a variation in velocity along the tank.

▲ **Figure 6** Here we zoom into the region near the front (left) in Figure 5, which should behave as if there are no sidewalls. (Thus, the thickness should not vary.) A black horizontal line is drawn on. Although this is a small length of shelf, the discrepancy with the sloped red line is obvious. Note that there are seeds in the Xanthan (for PIV).

We didn't see Figure 3 straight away, of course. It was quite something to see it slowly emerge – we didn't have the theory fully worked out until half the experiments were done. What was especially crucial was that we had the gradient measured before there was a theory, and the gradient told us how the frictional force scales with velocity. Knowing there was viscous drag in a narrow boundary layer near the walls, this measurement meant we realised within days what was going on, allowing the theory to be developed. This would have taken much longer without the data.

We also tracked particles that we put into the Xanthan. The results are shown below. The lateral velocity profile agreed very well with our model. The (peak) velocity where the PIV (particle tracking) was done is 0.246 ± 0.010 cm/s.

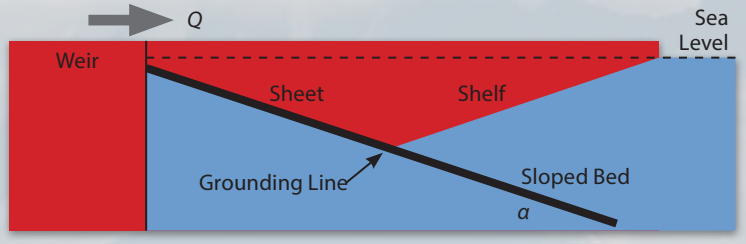
The speed of the front at the same time was 0.27 ± 0.01 cm/s. We expected the latter to exceed the former by 11%. The key thing is that such variations lead to forces other than from sidewalls, but such forces are not important for long enough shelves (we hoped). Proving that such velocity variations are small meant such forces were small and so our model was fundamentally correct (i.e. it got the dominant force balance). We also compared the thickness profile along the channel with our prediction (that it's nearly a perfect triangle). This also showed very good agreement.

We then turned our attention to better understanding ice tongues (i.e. without sidewalls). The above velocity profile would become uniform across the channel.

We expected the grounding line to stabilise, because like any floating object there should be some equilibrium. In our model, there's no tendency to thicken once some sort of equilibrium is attained. Based on the extended flatline at 1 on Figure 2 and our theory, we expected the front to go linearly with time for the whole experiment (no drag to slow it down)! The acceleration inherent in our theory (described previously) was predicted to be tiny in shelves of this length. Using the idea of forces along the channel balancing at the grounding line and using our understanding of what the forces are in the shelf (plus the viscous gravity current theory to determine this in the sheet), we created a computer model to predict the grounding line thickness.

We had previously done experiments without sidewalls. These were the most accurate ones we did, because we intended a 9 cm wide shelf in a 15 cm wide tank not to hit the sidewalls in a tank 90 cm long. The tank was levelled to within 2 arcminutes, so we succeeded. We quickly realised that the front speed was constant! But, how thick was the shelf? Of course, at constant Q and constant front speed (and constant width, see photograph) the thickness of the whole shelf had to be constant. Photographs (not shown) revealed no surprise. But, we hadn't designed the experiments to measure the grounding line thickness, which was quite low (1 cm at most).

In the end, we knew Q was constant and so was the front speed, and the width and height appeared unchanging in time and space. This allowed us to use flux conservation plus front speed and width



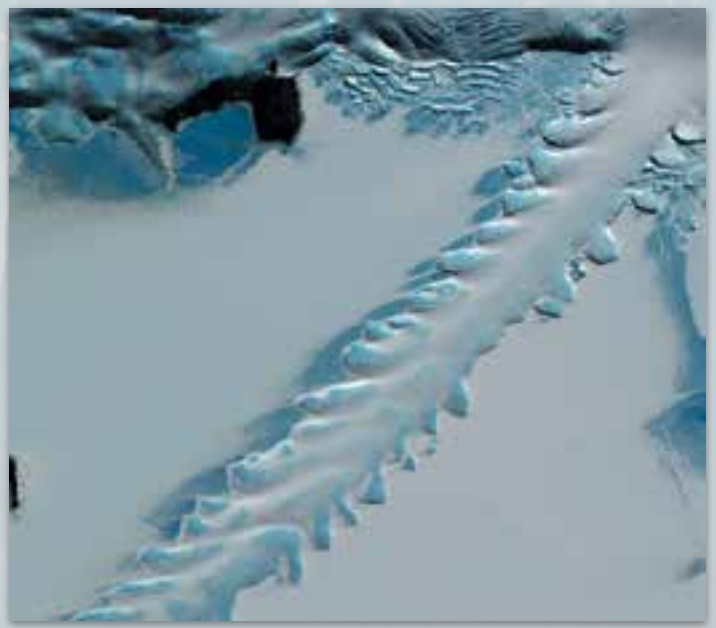
◀ **Figure 7** The situation when there is a grounding line. Note that the tapering of the shelf occurs (in a 1m tank) only if there is contact with the sidewalls.

measurements to infer the thickness of the shelf (we always knew Q extremely precisely). To check, we counted pixels in the side view, but the very low thickness and shadows etc. meant this was inaccurate. However, it revealed consistency with the indirect measurements outlined above. These agreed very closely with our computer model!

Next, we looked at an interesting effect (see comparison of real and laboratory models of ice tongues). This is due to a variable width at the grounding line, likely caused by a variable entry flux. After the grounding line, the shelf moves as a solid body. For a real glacier, Q oscillates annually but for us it oscillated due to the action of our pump (which is peristaltic). We still don't know precisely how Q affects the width, but the tapering at the front (when Q was rising from 0 to its final value, as the experiment had only just started); indicates that greater fluxes lead to a wider shelf.

The effort to understand more is still underway, but our involvement in it is likely over with the completion of this project. The end result is (this time) fewer questions than we started with, because some have definitely been answered, including the most crucial one – what's the dominant force balance (mathematically)?

If anyone wants to know what it took to get this far, basically it's determination and hard work. If we could see how to do something but it would be very hard, then we would always remind ourselves that we're lucky – sometimes, you can only wish you know what to do. Also crucial to our success was not feeling tied to anything that anybody (however experienced) predicted about the situation (i.e. believing it must be that way before the data came in). Instead of blindly trusting anybody's ideas, we believed in 'going with the flow', being guided by data and intuition and above all else having an unquenchable confidence in our ability to make at least partial progress on the road to understanding glaciers, even if nothing made sense and experiments didn't work (because who really knows what tomorrow will bring)?



Finding Order in Randomness

Maithra Raghu, Trinity

Given a complete graph on six vertices, denoted K_6 (a graph where every vertex is connected to every other vertex); we colour each edge of the graph either red or blue. Can we find a complete graph on three vertices (aka a triangle) such that all its edges are the same colour? What about for a bi-coloured K_{10}; can we find a monochromatic K_4?

In both cases, it is indeed possible. These problems are an example of the finite case of a theorem in Ramsey Theory.

Ramsey Theory is named after the British mathematician Frank Plumpton Ramsey (1903 – 1930) whose paper, *On a Problem of Formal Logic* (1928), proved what is now known as Ramsey's Theorem. This was not the first theorem proved in the area of modern Ramsey Theory; Issai Schur proved in 1916 that there always exist monochromatic x, y, z in a finite colouring of the naturals such that $x + y = z$ and Van der Waerden his eponymous theorem in 1927. However, Ramsey's work was imperative in ensuring that all these results were viewed collectively under one area and encouraging further research in this field.

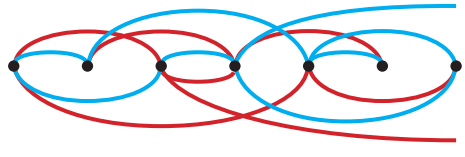
We have mentioned some problems for colouring a finite set of points; but what kind of patterns emerge if we colour an infinite set of points?

WNFC (When Naturals are Finitely Coloured)

We start simply, assuming as before, that we only have two colours, blue and red, at our disposal.

We now consider colouring edges of the complete graph with $\mathbb{N}$ as its vertex set with these two colours. This is a daunting thing to imagine, so we shall introduce some (abuse of) notation to make things easier.

Let $\mathbb{N}^{(2)}$ denote distinct pairs of natural numbers such that order does not matter i.e $(a, b) = (b, a)$. Then our colouring is simply the function $c : \mathbb{N}^{(2)} \to \{\text{blue, red}\}$.



We can now picture our daunting infinite complete graph as below, a sequence with pairs of the sequence connected by either red or blue lines. So a monochromatic subgraph in this context is simply a subset of $\mathbb{N}$ on which c is constant.

We now claim that we can find a subset M of $\mathbb{N}$, with

1. M infinite
2. $M^{(2)}$ monochromatic

To prove this, we pick any natural number a_1. All the lines coming out of a_1 are either red or blue. So there is some infinite subset B_1 of $\mathbb{N} - \{a_1\}$ such that all the lines from a_1 to B_1 are of the same colour (pigeonhole two and $\aleph_0$). We then pick a_2 in B_1 and pick infinite B_2 such that all lines from a_2 to B_2 are monochromatic. We keep repeating to get a sequence $a_1, a_2, a_3, \ldots$ with each a_i having either red or blue associated with it. As there are

only two colours, some colour occurs infinitely often and the terms associated with this colour give us our required set.

An interesting diversion

We can now prove that any sequence in a totally ordered set has a monotone subsequence. Let $\mathcal{C}(a_i, a_j)$ denote the colour of the edge (a_i, a_j). Then

1. $\mathcal{C}(a_i, a_j)$ = red if $a_i \geq a_j$;
2. $\mathcal{C}(a_i, a_j)$ = blue if $a_i < a_j$.

But now we know we have a monochromatic set, which corresponds to a monotone sequence.

This gives another way to prove the Bolzano-Weierstrass theorem! We let our totally ordered set be the reals, and we have proven that every bounded sequence has a monotone subsequence. But by the Fundamental Theorem for reals, any monotone sequence converges, so we have our convergent subsequence.

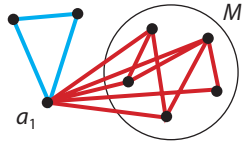
And back to where we left…

Two natural extensions follow:

1. When two-colouring $\mathbb{N}^{(r)}$ for some finite r, can we find a subset M as before?
2. What happens if we use some finite number k of colours instead?

It is indeed possible to find a set M as before in both cases.

For part 1, we proceed by induction. The case $r = 1$ is an application of the Pigeonhole principle, and the case $r = 2$ is what we have just proven. So let us assume the result for $r = k$ and consider $r = k + 1$. As before, we pick some a_1 in $\mathbb{N}$. Now notice that this induces a two colouring on $(\mathbb{N} - \{a_1\})^{(r-1)}$. The diagram shows the case for $r = 3$:



And by our induction hypothesis, there is an infinite monochromatic set M for this colouring, say colour red. But as the colouring was induced on our $(r-1)$-tuples in M by removing, when we add $\{a_1\}$ back; we recover our r-tuple colouring, which has the same colouring as the $(r-1)$-tuple.


Frank Plumpton Ramsey (1903 – 1930)

So $M \cup \{a_1\}$ provides the set we were looking for.

And now for part 2, increasing the number of colours we can create disorder with. The case for colouring $\mathbb{N}^{(2)}$ with some finite number k of colours turns out to be surprisingly easy, using the ideas of induction and "colour-blindness".

We induct on the number of colours k. We assume that we only have two colours, red and everything-but-red. Then by our previous work, we know there is a monochromatic set M. If M is red, then we are done! If not, then we've just reduced to $k-1$ colours, which is soluble by the induction hypothesis.

Infinite to Finite

Before plunging back into the infinite, note that we can prove finite Ramsey from what we know of infinite Ramsey. We write $[n] = \{1, 2, 3, \ldots, n\}$.

> **Theorem** Let $m, r \in \mathbb{N}$. Then there exists n such that whenever $[n]^r$ is two coloured, there is a monochromatic set $A \subset [n]$ of size m.

The proof is left as an exercise. Try and construct a two colouring of $\mathbb{N}^{(r)}$ without a monochromatic M, providing the contradiction.

The canonical Ramsey Theorem

We have managed to avoid the most daunting question yet. What happens if we dare to colour $\mathbb{N}^2$ with infinitely many colours? The question seems inane; what kind of pattern could we hope

▲ Possible Structures of M

to find? A monochromatic set M is out of the question; we have the power to pathologically colour every point in $\mathbb{N}^2$ a different colour. Yet even in this melange of colourings there is some order.

Theorem If we have an arbitrary colouring of $\mathbb{N}^{(2)}$, there exists an infinite M of one of the following forms, for arbitrary $i < j < k < l$:

(i) $M^{(2)}$ is monochromatic;

(ii) Each point in $M^{(2)}$ has a different colour;

(iii) (i, j), (k, l) in $M^{(2)}$ have the same colour iff $i = k$;

(iv) (i, j), (k, l) in $M^{(2)}$ have the same colour iff $j = l$;

These four possibilities are shown in the diagrams above.

Here we give a sketch proof of this theorem. In the proof, we are going to work with sets of the form $A^{(4)}$ and colour with the colours of 'same' and 'different' (based on certain properties of our set), and then use our previous work to find sets monochromatic in 'same' or 'different' to filter out our desired properties.

We first deal with the monochromatic case.

First we two-colour $\mathbb{N}^4$ by giving (i, j, k, l) colour

- 'same' if $\mathcal{C}(i, j) = \mathcal{C}(k, l)$ in original colouring;
- 'different' otherwise.

By our previous work, there exists an infinite monochromatic set A_1. If A_1 is the colour 'same' then it is monochromatic (exercise). So we assume A_1 is 'different'.

Note that A_1 being 'different' implies that $\mathcal{C}(i, j)$ never equals $\mathcal{C}(j, k)$ in A_1, else we obtain a contradiction (exercise).

We define 'left same' to be if $\mathcal{C}(i, j) = \mathcal{C}(i, k)$ and 'left different' otherwise, and 'right same' as $\mathcal{C}(j, k) = \mathcal{C}(i, k)$ and 'right different' analogously. Now we two colour A_1 twice, to obtain A_2 and A_3 as follows. We first find an A_2 monochromatic in 'left same' / 'left different' and from that form a monochromatic A_3 in 'right same' / 'right different'. We now have various different cases:

- If A_3 is 'left different' and 'right same' then it is case (iv) from above.
- If A_4 is 'left same' and 'right different' then it is case (iii) from above.
- Note that A_3 cannot be right and left same, as it is a contradiction to A_1 being 'different'.
- Finally, we have the case where A_3 is different in both sides. Note that we can find a subset A_5 such that A_5 is monochromatically different both regarding $\mathcal{C}(i, l) = \mathcal{C}(j, k)$ as 'same' and $\mathcal{C}(i, k) = \mathcal{C}(j, l)$ as 'same'.

In each case, we can find a monochromatic 'different' set M, as M being 'same' would result in a contradiction with A_1 being 'different'. Indeed, if we found a 'same' M for $\mathcal{C}(i, l) = \mathcal{C}(j, k)$, then picking $i < j < k < l < m < n$, we have $\mathcal{C}(i, n) = \mathcal{C}(j, k)$ and $\mathcal{C}(i, n) = \mathcal{C}(l, m)$, so $\mathcal{C}(j, k) = \mathcal{C}(l, ,)$ in A_1, which is a contradiction.

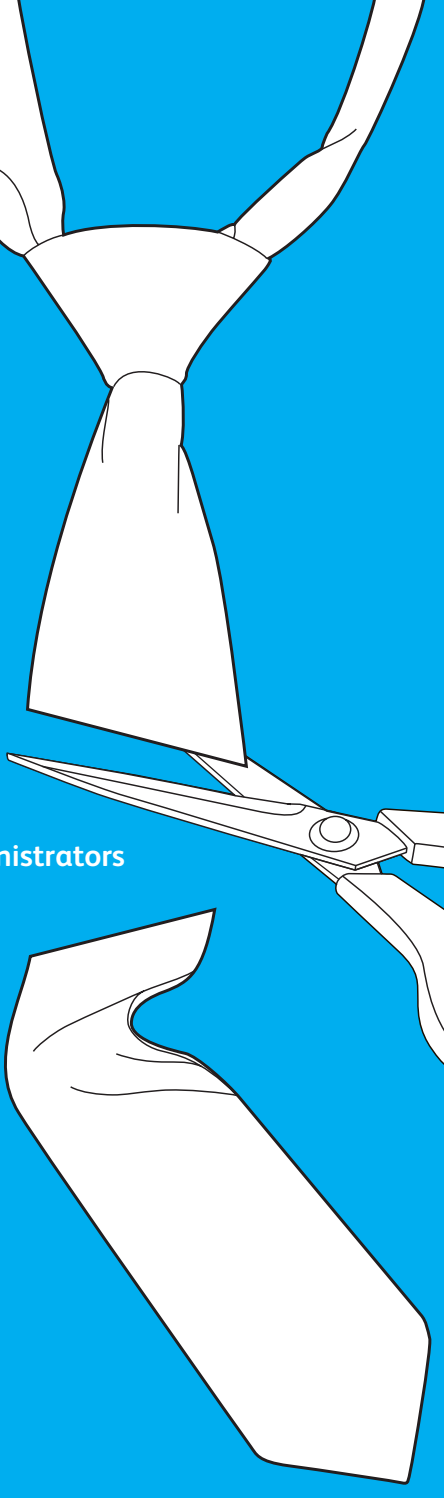
This A_2 satisfies case (ii) from above, and we are done!

In Conclusion

In this article, we have focussed solely on Ramsey's Theorem. But modern Ramsey Theory extends far beyond this. Van der Waerden's theorem looks at finding monochromatic arithmetic progressions. Though originally considered a cornerstone of Ramsey Theory, Richard Rado's astounding extension of Schur's Theorem, suitably called Rado's Theorem, gives an immediate solution to Van der Waerden and all its extensions. Hindman's Theorem (proved 1973) then extends Rado's ideas of partition regularity to an infinite setting. Despite these leaps in understanding, we still lack basic information in many areas of Ramsey Theory. Indeed, even now, it is not entirely clear how Rado managed to conceive his ground breaking theorem. This absence of full understanding, combined with the relative clarity of the problems ensures that Ramsey Theory will be a fruitful and fascinating area of research in the future.

G-RESEARCH

WE OFFER INTELLECTUAL FREEDOM NOT CORPORATE CONSTRAINTS



Quantitative Researchers | Developers | Systems Administrators

The best minds thrive on freedom. Take our teams; they're always striving to produce groundbreaking research, software and IT systems for use in investment management. Because we can offer them a flexible, informal working culture, they succeed. After all, intellects like theirs can't be constrained. What about yours? **www.gresearch.co.uk.**

The next level

Latest technologies
Mathematical challenges
Real world problems

Mathematics in Wartime

G H Hardy

First published in issue 3, 1940

The editor asked me at the beginning of term to write an article for EUREKA, and I felt that I ought to accept the invitation; but all the subjects which he suggested seemed to me at the time quite impossible. "My views about the Tripos" – I have never really been much interested in the Tripos since I was an undergraduate, and I am less interested in it now than ever before. "My reminiscences of Cambridge" – surely I have not yet come to that. Or, as he put it, "something more topical, something about mathematics and the war" – and that seemed to me the most impossible subject of all. I seemed to have nothing at all to say about the functions of mathematics in war, except that they filled me with intellectual contempt and moral disgust.

I have changed my mind on second thoughts, and I select the subject which seemed to me originally the worst. Mathematics, even my sort of mathematics, has its "uses" in war-time, and I suppose that I ought to have something to say about them; and if my opinions are incoherent or controversial, then perhaps so much the better, since other mathematicians may be led to reply.

I had better say at once that by "mathematics" I mean *real* mathematics, the mathematics of Fermat and Euler and Gauss and Abel, and not the stuff which passes for mathematics in an engineering laboratory. I am not thinking only of "pure" mathematics (though that is naturally my first concern); I count Maxwell and Einstein and Eddington and Dirac among "real" mathematicians. I am including the whole body of mathematical knowledge which has permanent aesthetic value, as for example, the best Greek mathematics has, the mathematics which is eternal because the best of it may, like the best literature, continue to cause intense emotional satisfaction to thousands of people after thousands of years. But I am not concerned with ballistics or aerodynamics, or any of the other mathematics which has been specially devised for war. That (whatever one may think of its purposes) is repulsively ugly and intolerably dull; even Littlewood could not make ballistics respectable, and if he could not, who can?

Let us try then for a moment to dismiss these sinister by-products of mathematics and to fix our attention on the real thing. We have to consider whether real mathematics serves any purposes of importance in war, and whether any purposes which it serves are good or bad. Ought we to be glad or sorry, proud or ashamed, in war-time, that we are mathematicians?

It is plain at any rate that the real mathematics (apart from the elements) has no *direct* utility in war. No one has yet found any war-like purpose to be served by the theory of numbers or relativity or quantum mechanics, and it seems very unlikely that anybody will do so for many years. And of that I am glad, but in saying so I may possibly encourage a misconception.

It is sometimes suggested that pure mathematicians glory in the "uselessness" of their subject, and make it a boast that it has no "practical" applications. I have been accused of taking this view

Days needed by Phileas Fogg to travel around the world. Number of columns in IBM punch cards.

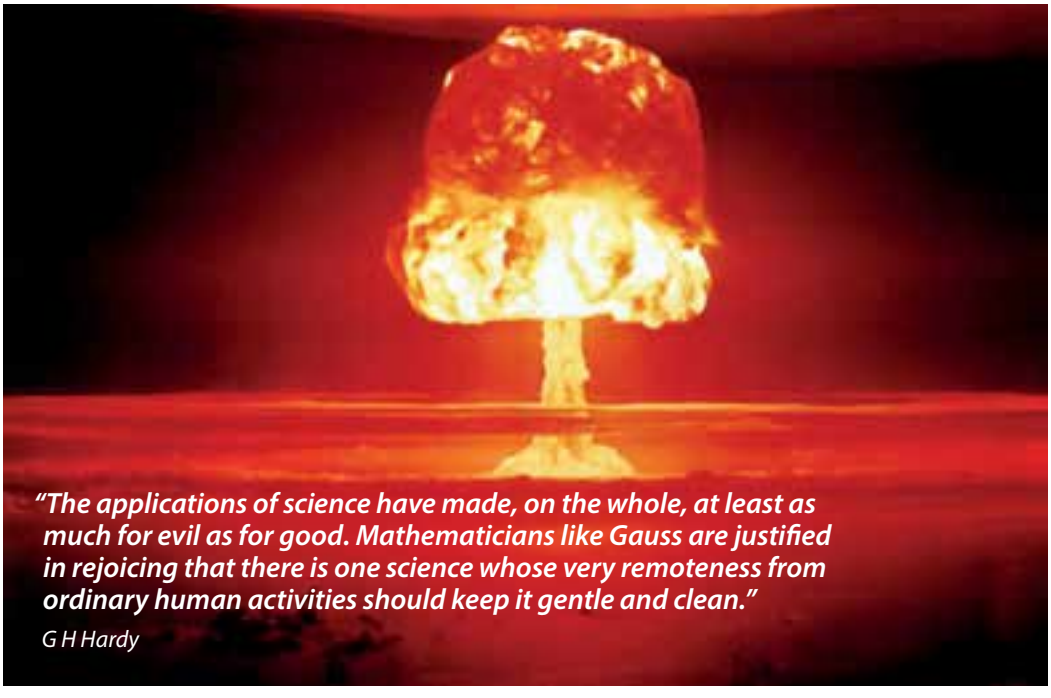
Godfrey Hardy (1877 – 1947) is not only famous ▶ for his achievements in number theory and analysis, but for mentoring Srinivasa Ramanujan and for writing *"A Mathematician's Apology"* on the aesthetics of mathematics.

myself. I once stated in a lecture, which was afterwards printed, that "a science is said to be useful if its development tends to accentuate the existing inequalities in the distribution of wealth, or more directly promotes the destruction of human life"; and this sentence, written in 1915, was quoted in the *Observer* in 1939. It was, of course, a conscious rhetorical fluorish (though one perhaps excusable at the time when it was written).

The imputation is usually based on an incautious saying attributed to Gauss, to the effect that, if mathematics is the queen of the sciences, then the theory of numbers is, because of its supreme "uselessness," the queen of mathematics, which has always seemed to me to have been rather crudely misinterpreted. If the theory of numbers could be employed for any practical and honourable purpose, if it could be turned directly to the furtherance of human happiness or the relief of human suffering (as for example physiology and even chemistry can), then surely neither Gauss nor any other mathematician would have been so foolish as to decry or regret such applications. But if on the other hand the applications of science have made, on the whole, at least as much for evil as for good – and this is a view which must always be taken seriously, and most of all in time of war – then both Gauss and lesser mathematicians are justified in rejoicing that there is one science at any rate whose very remoteness from ordinary human activities should keep it gentle and clean.

It would be pleasant to think that this was the end of the matter, but we cannot get away from the mathematics of the workshops so easily. Indirectly, we are responsible for its existence. The gunnery experts and aeroplane designers could not do their job without quite a lot of mathematical training, and the best mathematical training is training in real mathematics. In this indirect way even the best mathematicians becomes important in war-time, and mathematics are wanted for all sorts of purposes. Most of these purposes are ignoble and dreary – what could be more soul-destroying than the numerical solution of differential equations? – but the men chosen for them must be mathematicians and not laboratory hacks, if only because they are better trained and have the better brains. So mathematics is going to be really important now, whether we like it or regret it; and it is not so obvious as it might seem at first even that we ought to regret it, since that depends upon our general view of the effect of science on war.

There are two sharply contrasted views about modern "scientific" war. The first and the most obvious is that the effect of science on war is merely to magnify its horror, both by increasing the sufferings of the minority who have to fight and by extending them to other classes. This is the orthodox view, and it is plain that, if this view is just, then the only possible defence lies in the necessity for retaliation. But there is a very different view which is also quite tenable. It can be maintained that modern warfare is *less* horrible than the warfare of pre-scientific times, so far at any rate as combatants are concerned; that bombs are probably more merciful than bayonets; that lachrymatory gas and mustard-gas are perhaps the most humane weapons yet devised by military



science, and that the "orthodox" view rests solely on loose-thinking sentimentalism. This is the case presented with so much force by Haldane in *Callinicus*. It may also be urged that the equalisation of risks which science was expected to bring would be in the long run salutary; that a civilian's life is not worth more than a soldier's, or a woman's than a man's; that anything is better than the concentration of savagery on one particular class; and that, in short, the sooner war comes "all out" the better. And if this be the right view, then scientists in general and mathematicians in particular may have a little less cause to be ashamed of their profession.

It is very difficult to strike a balance between these extreme opinions, and I will not try to do so. I will end by pulling to myself, as I think every mathematician ought to, what is perhaps an easier question. Are there *any* senses in which we can say, with any real confidence, that mathematics "does good" in war? I think I can see two (though I cannot pretend that I extract a great deal of comfort from them).

In the first place it is very probable that mathematics will save the lives of a certain number of young mathematicians, since their technical skill will be applied to "useful" purposes and will keep them from the front. "Conservation of ability" is one of the official slogans; "ability" means, in practice, mathematical, physical, or chemical ability; and if a few mathematicians are "conserved" then that is at any rate something gained. It may be a bit hard on the classics and historians and philosophers, whose chances of death are that little much increased; but nobody is going to worry about the "humanities" now. It is better that some should be saved, even if they are not necessarily the most worthy.

Secondly, an older man may (if he not *too* old) find in mathematics an incomparable anodyne. For mathematics is, of all the arts and sciences, the most austere and the most remote, and a mathematician should be of all men the one who can most easily take refuge where, as Bertrand Russell says, "one at least of our nobler impulses can best escape from the dreary exile of the actual world." But he must not be too old – it is a pity that it should be necessary to make this very serious reservation. Mathematics is not a contemplative but a creative subject; no one can draw much consolation from it when he has lost the power or the desire to create; and that is apt to happen to a mathematician rather soon. It is a pity, but in that case he does not matter a great deal anyhow, and it would be silly to bother about him.

The sixth nuclear magic number. Number of 6-hexes (shapes made by joining six regular hexagons).

Mitered Fractal Tree

Koos Verhoeff and Anton Bakker

'Best of Show' at the 2012 Bridges Conference on Mathematics and Art

Consecutive Integers

Paul Erdős

First published in issue 38, 1975/76

Some time ago, two old problems on consecutive integers were settled. Catalan conjectured that 8 and 9 are the only consecutive powers. First of all observe that four consecutive integers cannot all be powers since one of them is congruent to 2 modulo 4.

It is considerably more difficult to show that three consecutive integers can not all be powers; this was accomplished about 20 years ago by Cassels and Makowski. Finally in 1974 using some deep results of Baker, Tijdeman proved that there is an n_0, whose value can be given explicitly, such that for $n > n_0$, n and $n+1$ are not both powers. This settles Catalan's conjecture almost completely.

It has been conjectured that if $x_1 < x_2 < x_3 < \dots$ is a sequence of consecutive powers, such as $x_1 = 1$, $x_2 = 4$, … then $x_{i+1} - x_i > i^c$ for all i and some suitable constant c. At the moment this seems intractable.

It was conjectured more than a century ago that the product of consecutive integers is never a power. Almost 40 years ago, Rigge and I proved that the product of consecutive integers is never a square, and recently Selfridge and I proved the general conjecture. In fact, our result is that for every k and l there exists a prime $p \geq k$ such that if

$$p^{\alpha_{k,l}} \mid \prod_{i=1}^{k} (n+i)$$

then

$$a_{k,l} \equiv 1 \bmod p.$$

We conjecture that in fact for all $k > 2$ there is a prime $p \geq k$ with $a_{k,l} \equiv 1$, but this is also intractable at the moment.

It often happens in number theory that every new result suggests many new questions – which is a good thing as it ensures that the supply of Mathematics is inexhaustible! I would now turn to discuss a few more problems and results on consecutive integers and in particular a simple conjecture of mine which is more than 25 years old.

Put

$$m = a_k(m)\, b_k(m),$$
$$a_k(m) = \prod p^{\alpha_p},$$

where the product extends over all the primes $p \geq k$ and $p^{\alpha} \mid m$. Further define

$$f(n; k, l) = \min\{a_k(n+i) : 1 \leq i \leq l\};$$

$$F(k, l) = \min\{f(n; k, l) : 1 \leq n \leq \infty\}.$$

I conjectured that

$$\lim_{k\to\infty} \frac{F(k,k)}{k} = 0. \qquad (1)$$

In other words, is it true that for every ε there is a k_ε such that for every $k > k_\varepsilon$ at least one of the integers $a_l(n+i)$ for $i = 1, \dots, l$, is less than k_ε. I am unable to prove this but will outline the proof of

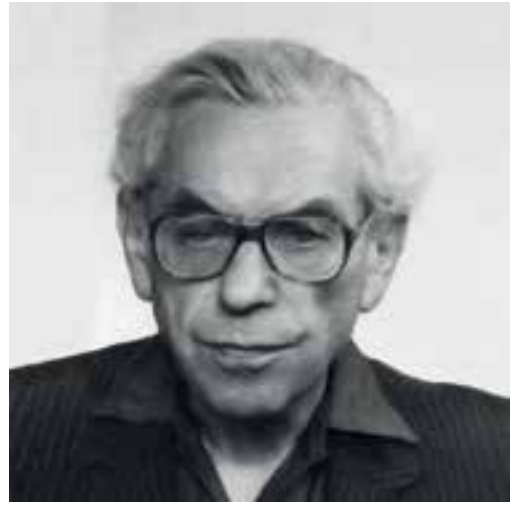
$$F(k,k) < (1+\varepsilon)\, k \quad \text{for} \quad k > k_0(\varepsilon). \qquad (2)$$

To prove (2) consider

$$A(n,k) = \prod\nolimits'_{i=1}^{k} a_l(n+i) \qquad (3)$$

Number of 5-faces of a seven-dimensional hypercube.
Largest order of a permutation of 14 elements.

Paul Erdős (1913 – 1996) has published more ▶ papers than any other mathematician in history.



where the $\prod'$ in (3) indicates that for every $p \le k$ we omit one of the integers $n+i$ divisible by a maximal power of p. Then the product $\prod' a_k(n+i)$ has at least $k - \pi(k)$ factors and by a simple application of the Legendre formula for the factorisation of $k!$ we obtain

$$\prod{}' a_k(n+i) \mid k! \tag{4}$$

If (2) did not hold, we have from (4) and Stirling's formula

$$((1+\varepsilon)k)^{k-\pi(k)} < k^{k+1}e^{-k} \tag{5}$$

or $$k^{\pi(k)+1} > e^k(1+\varepsilon)^{k-\pi(k)}.$$

Now, by the prime number theorem,

$$\pi(k) < \frac{(1+\varepsilon/10)k}{\log k},$$

and so from (5),

$$k + \left(\frac{(1+\varepsilon/10)k}{\log k} + 1\right) > (1+\varepsilon)e^k + \left(k - \frac{2k}{\log k}\right),$$

which is false if k is large enough. This contradiction proves (2).

Assume for the moment that (1) has been proved. Then one can immediately ask for the true order of magnitude of $F(k,k)$. I expect that it is $o(k^\varepsilon)$ for every $\varepsilon > 0$. On the other hand, I can prove that

$$F(k,k) > \exp\left(c \cdot \frac{\log(k)\log\log\log(k)}{\log\log(k)}\right) \tag{6}$$

The problem of estimating $F(k,k)$ and the proof of (6) is connected with the following question on the sieve of Eratosthenes-Prim-Selberg : determine or estimate the smallest integer $A(k)$ so that one can find, for every p with $A(k) \le p \le k$, a residue u_p such that for every integer $t \le k$, t satisfies one of the congruences to u_p modulo p. Clearly $F(k,k) \nless A(k)$. Using the method of Rankin-Chen and myself I proved

$$A(k) > \exp\left(c \cdot \frac{\log(k)\log\log\log(k)}{\log(k)}\right) \tag{7}$$

which implies 6. I do not give the proofs here. It would be interesting and useful to prove $A(k) < k^\varepsilon$ for every $\varepsilon > 0$ and sufficiently large k.

Now, I shall say a few words about $F(k,1)$ for $k \ne 1$.

It follows easily from the Chinese Remainder Theorem that for $1 \le \pi(k)$ we have $F(k,l) = \infty$, since for a suitable n, we can make $n+i$ for $1 \le i \le \pi(k)$ divisible by an arbitrarily large power of p_1. It is easy to see that this no longer holds for $l = \pi(k)+1$ and in fact it is not hard to prove that

$$F(k, \pi(k)+1) = \prod p^{\alpha_p},$$

where $p^{\alpha_p} \le \pi(k) < p^{\alpha_p+1}$. As l increases it gets much harder to even estimate $F(k,l)$.

Many more problems can be formulated which I leave to the reader and only state one which is quite fundamental: Determine or estimate the least $l = l_k$ so that $F(k,l_k) = 1$.

In other words, the least l_k so that among l_k consecutive integers there is always one relatively prime to the primes less than k. This question is of course connected with the problem of estimat-

ing the difference of consecutive primes and also with the following problem of Jacobsthal: Denote by $g(m)$ the least integer so that any set of $g(m)$ consecutive integers contains one which is relatively prime to m. At a Number Theory meeting in Oberwolfach (November '75), Kanold gave an interesting talk on $g(m)$. Vaughan observed that the sieve of Rosser gives $g(m) < (\log m)^{2+\varepsilon}$ for all $\varepsilon > 0$ if m is sufficiently large. The true order of magnitude is not known.

It seems to me that interesting and difficult problems remain for $1 \le \pi(k)$ too. Here we have to consider the dependence on n too. It is not hard to show that for every $\varepsilon > 0$ there are infinitely many values of n for which

$$f(n;k,l) > (1-\varepsilon)^{1/n}. \tag{8}$$

The proof of (8) uses some elementary facts of Diophantine approximation and the Chinese Remainder Theorem. We do not give the details. I do not know how much (8) can be improved. By a deep theorem of Mahler, using the p-adic Thue-Siegel Theorem, $f(n;k,l) > n^{\varepsilon+1/l}$. It is quite possible that

$$\lim_{n\to\infty} \sup \frac{f(n;k,l)}{n} = \infty. \tag{9}$$

Interesting problems can also be raised if k tends to infinity with n; e.g. how large can $f(n;k,\pi(k))$ become if $k = (1+o(1))\log n$? It seems to be difficult to write a really short note on the subject since new problems occur while one is writing!

It would be of some interest to know how many of the integers $a_k(n+i)$ must be different. I expect that more than $c \times k$ are. If this is proved one of course must determine the best possible value of c.

Denote by $K(l)$ the greatest integer below l composed entirely of primes below k. Trivially

$$\min_n \max_i a_k(n+i) = K(l). \tag{10}$$

To prove (10) observe that on the one hand any set of l consecutive integers contains a multiple of $K(l)$ on the other that if $2l$ divides t, then the integers $t!+1, \ldots, t!+l$ clearly satisfy (10), when $n = 0$. More generally, try to characterise the set of n which satisfy (10). To simplify matters, let $k = 1$ and denote n_k as the smallest positive integer with $\max_i a_k(n+i) = k$, S_k as the class of all integers n such that this is true. If p^{a_p} is the greatest power of p not exceeding k then

$$\prod_{p\le k} p^{a_p+1} \in S_k.$$

Perhaps I am overlooking an obvious explicit construction for n_k but at the moment I do not even have good upper or lower bounds for it. When is $k!$ in S_k? The smallest such k is 8 and I do not know if there are infinitely many such k's. But by Wilson's theorem, $p!$ is never in S_p.

To complete this note, I state three more extremal problems in number theory. Put

$$n! = \prod a_i, \text{ for } a_1 \le a_2 \le \ldots \le a_n.$$

Determine $\max a_1$. It follows easily from Stirling's formula that a_1 does not exceed $\frac{n}{e}\left(1-\frac{c}{\log n}\right)$. I conjectured that for every $\eta > 0$ and sufficiently large n, $\max a_1$ exceeds $\frac{n}{e}(1-\eta)$.

Now write

$$n! = \prod b_i, \text{ for } 1 < b_1 < b_2 < \ldots < b_n \le n.$$

Determine or estimate $\min k$.

Clearly k exceeds $n - n/\log n$ and by more complicated methods I can prove

$$k = n - n(1+o(1))/\log n,$$
$$k > n - n(\log n + c)/(\log n)^2,$$

where c is a positive absolute constant.

Finally write

$$n! = \prod u_i, \text{ for } u_1 < u_2 < \ldots < u_k. \tag{11}$$

Determine or estimate $\min u_k$, but k is not fixed. It is not hard to prove that u_k less than $2n$ has only a finite number of solutions. I only know of two:

$$6! = 8 \times 9 \times 10,$$
$$14! = 16 \times 21 \times 22 \times 24 \times 25 \times 26 \times 27 \times 28.$$

It would be difficult to determine all the solutions, although Vaughan has found some more:

$$3! = 6,$$
$$8! = 12 \times 14 \times 15 \times 16,$$
$$15! = 16\times18\times20\times21\times22\times25\times26\times27\times28,$$

and these are all up to 15. Vaughan also tells me

$$40! = 42\times44\times45\times48\times49\times50\times51\times52\times 54\times55\times56\times57\times58\times59\times60\times62\times63\times 64\times65\times66\times68\times69\times72\times74\times80.$$

Can be written as 222 in base 6. Number of metals in the periodic table. Happy number.

The Ultimate Painting

Drop Artists, 1966

Archimedes

Tom Körner, DPMMS Cambridge

Archimedes was the greatest mathematician, possibly the greatest scientist and certainly one of the greatest engineers of antiquity. Plutarch writes that although his engineering achievements '*gave him the renown of more than human sagacity, … he placed his whole affection and ambition in those purer speculations where there can be no reference to the vulgar needs of life; studies, the superiority of which to all others is unquestioned, and in which the only doubt can be, whether the beauty and grandeur of the subjects examined, or the precision and cogency of the methods and means of proof, most deserve our admiration.*'

Archimedes' final triumph as an engineer was the defence of Syracuse (in 212 BC) when '*such terror seized the Romans, that, if they did but see a little rope or a piece of wood from the wall, instantly crying out, that there it was again, Archimedes was about to let fly some engine at them, they turned their backs and fled*'. However, the Romans eventually prevailed and he died in the sack of the city. The Romans, who organised the destruction of more people of every race, religion and colour than any empire before, were always a little ashamed of killing the greatest mind of antiquity and invented several fine stories about his death.

The writings of Archimedes were collected, copied and expounded for the next 1500 years but, although some of those who studied them certainly understood them, they do not seem to have progressed much beyond him. One reason for this may have been expressed by Cicero, who was proud of restoring the tomb of the great man. '*Among them [the Greeks] geometry was held in highest honour; nothing was more glorious than mathematics. But we [the Romans] have limited the usefulness of this art to measuring and calculating.*' (Or, as EPSRC might put it, '*shaping capability*'.) It is perhaps, not surprising that, although the Romans produced much bigger and somewhat better versions of existing technologies, they produced little that was entirely novel. Another reason for the lack of progress may have been the channelling of Greek abstract thought into the of endless marshes of Christian theological controversy.

The fall of the Eastern Roman Empire, ending with the sack of Constantinople in 1492, resulted in the loss of an enormous number of Greek manuscripts. Which books survived seems to have been mainly a matter of luck. Some of Archimedes' works survived as earlier translations into Arabic, but most of what survived was in two manuscripts which found their way into the possession of the Norman kings of the Two Sicily's and then into the Vatican library. Both have since disappeared, but not before they were translated into Latin first by William of Markab in 1296 and then by James of Crewman in 1544.

The invention of printing meant that Crewman's translation could be widely distributed and Archimedes became a hero and a source of inspiration to early scientists like Kepler and Galileo. Archimedes showed them that mathematics could be used not merely to study the heavens (which had always had an ethereal and so mathematical feel) but everyday things like boats floating in water. Newton wrote his Principia in the Greek (that is

to say, the Archimedian) style, making a difficult book even more difficult, but showing the respect due from one mage to another.

The new calculus of Newton and Leibniz meant that any fool (measured on the Newton and Archimedes scale) could find results which up then had required the genius of Archimedes and his direct influence on science came to an end.

In 1906, a Danish classical scholar named Heiberg realised that a prayer book held in Istanbul was written on reused vellum and that the original text (which had, of course, been carefully scraped off) was a collection of Archimedean works. Working from photographs, he was able to recover most of this text. Several of the works were known in one form or another but one now called *The Method of Mechanical Theorems* created a sensation.

It begins with words that still thrill me many years after I first read them: *'Archimedes to Eratosthenes greeting. […] Seeing moreover in you, as I say, an earnest student, a man of considerable eminence in philosophy, and an admirer [of mathematical enquiry], I thought fit to write out for you [...] the peculiarity of a certain method, by which it will be possible [...] to investigate some of the problems in mathematics by means of mechanics. This procedure is, I am persuaded, no less useful even for the proof of the theorems themselves; for certain things first became clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards because their investigation by the said method did not furnish an actual demonstration.'*

In other words, the great magician will draw back the curtain and reveal his secrets. And those secrets turn out to be tremendous – not quite the modern calculus of Newton and Leibniz but certainly containing many of the ideas, painfully discovered by their predecessors, which underlie that calculus. Alternative history is a mug's game, but it is hard not to feel that, if Galileo or Kepler had held The Method in their hands, Western science would have been advanced by fifty years. Archimedes concluded his introduction with the words. *'I am persuaded that [this method] will be of no little service to mathematics; for I apprehend that some, either of my contemporaries or of my successors, will, by means of the method when once established, be able to discover other theorems in addition, which have not yet occurred to me.'* But it was not to be.

Naturally scholars returned to Istanbul to look at the original prayerbook, only to find it had disappeared!

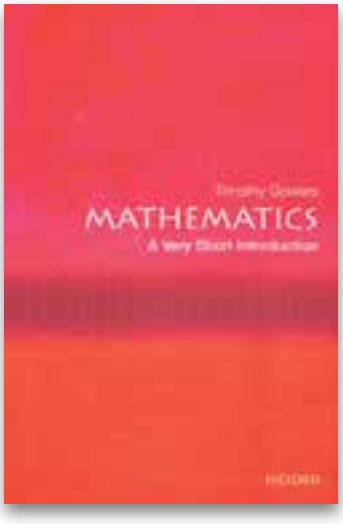
The Archimedes Palimpsest
ISBN-13: 9781107014572
$140.00

In 1998 it reappeared, further damaged, in part, by neglect and, in part, by a criminal attempt at forgery, and was sold at auction to an anonymous American for a mere two million dollars. Fortunately it was now in the charge of someone who knew its true value. The most modern scientific techniques have been used to study it and the results are now issued in two beautiful volumes by CUP. (More technical and scholarly volumes will follow.) The first volume gives the background to the studies and the second images of the restoration itself.

From the point of view of the mathematician, little more is revealed than was known through the work of Heiberg. However classicists were thrilled by the discovery of several speeches of Hypereides (one of the major Greek orators), a commentary on Aristotle and several as yet unidentified fragments.

It is unlikely that many people will fork out 150 pounds to buy these two volumes, particularly since one of them is in ancient Greek. But those who do will own a triumph of the art of making books, a triumph of the ability of modern science to make darkness visible, a triumph in the classicist's six hundred year struggle to to restore the wisdom of the ancients and a monument to a man who more than two thousand years ago helped lay the foundations of the modern world. 'History is indeed little more than the register of the crimes, follies and misfortunes of mankind' but occasionally we get a glimpse of something better. These volumes are, as it were, the concentrated essence of civilisation.

Book Reviews



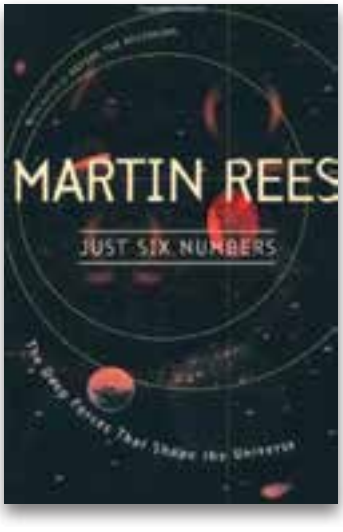
Mathematics: A Very Short Introduction

Timothy Gowers — *Oxford University Press, 2002*
ISBN 978-0-19-285361-5 — *£7.99*

A thoroughly entertaining little book that lends itself well to casual reading, and which justifies its title wonderfully. While mathematics students may find the concepts rather familiar or basic, Gowers' lucid style and simple examples make the content accessible to all.

Encouraging the reader to think abstractly, the book touches on topics such as fractional dimension, hyperbolic geometry and uses of mathematical models. Its essentially independent chapters can be read separately, but at the same time are neatly unified by the underlying philosophical flavour. Some may also find a Fields Medallist's responses to oft-asked questions including *"Is it true that mathematicians are past it by the time they are 30?"* and *"Why are there so few women mathematicians?"* intriguing, in the last chapter.

This is an insightful bridge between the mathematics taught at school and what aspiring students can look forward to, and is recommended for anyone with an interest in the subject. *Stacey Law*

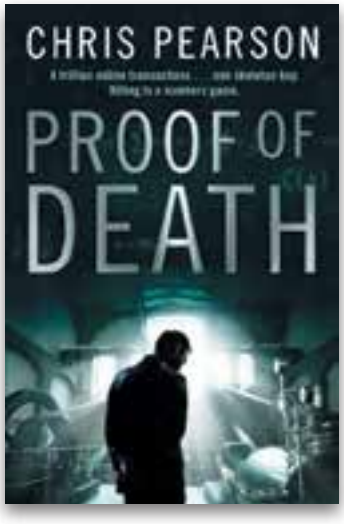


Just Six Numbers

Sir Martin Rees — *Phoenix, 2001*
ISBN: 978-0-75-381022-4 — *£8.99*

Disregarding how abstract the topic is, a good mathematics book should be understood at some level by any reader. When I started my undergraduate course I just understood the basics from this book, that there are six main numbers that define cosmology: the number of dimensions we live in, the ratio of the strength of gravity to that of electromagnetism, ε, the ratio of mass lost to energy when hydrogen is fused to form helium, Ω, describing the amount of dark matter, λ, the cosmological constant, and Q, related to the scale at which the universe looks smooth.

In time I understood the rest of the book. It is a really good book to start with, since Martin Reese has managed to explain the key ideas behind cosmology today in 180 pages without any "fuss" equations. And it proves that cosmology can be done while having a nice cup of tea. *Carina Negreanu*

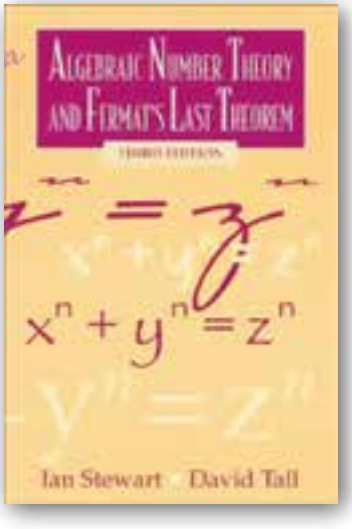


Proof of Death

Chris Pearson — *Kindle eBook*
ASIN: B008U8R20K — *£0.99*

From a grim scene of hostages in Chechnya to the Great Clock in Trinity College, Pearson creates a web of mystery around the fictional proof of the Riemann Hypothesis. Weaving together various locations and plots to keep you avidly reading to the end, his thriller cleverly incorporates both number theory and its application to cryptography. The plot is skilfully designed so that mathematicians and non-mathematicians alike are sucked into Aslan's world of survival and deceit.

"A prime number, of course – no divisors except itself and one – always yielded the best brew." Packed with emotion and description, this book is sure to provide a fictional world of mathematical proof that any reader can easily delve into. *Eleanor Wale, Reading*



Algebraic Number Theory and Fermat's Last Theorem

Ian Stewart, David Tall — *AK Peters, 2002*
ISBN 1-56881-119-5 — *£37.99*

It is difficult to find a mathematics book that is both precise and informal. This book has both qualities, giving historical background information while rigorously developing algebraic number theory. It is suitable for undergraduates meeting the subject for the first time. Definitions are motivated and important concepts are illustrated by computational examples.

The material in the first 10 chapters is approximately equivalent to the Part II Number Fields course, landmarks being ideals, Minkowski's Theorem, and the class-group. The remaining chapters contain the proof of a special case of Fermat's Last Theorem (regular prime exponents), which uses all the previously introduced ideas. They also touch on more advanced topics leading up to a sketch proof of its general version.

The extra material on elliptic curves and elliptic functions has little to do with the rest of the book and feels a bit disconnected. However, the chapters on algebraic number theory are excellent for accompanying a university course, while the last part will whet the reader's appetite for more. *Philipp Kleppmann*

Archimedeans
Christmas Catalogue

Klein Salad Dressing Bottle

Keep oil on the inside, vinegar on the outside

£20

Random Walk Generator

Comprises stereotypical mathematician and half-pint of larger.

FREE

3D Random Walk Generator

Also includes centrifuge and trampoline.

Accessories: reflecting barriers, absorbing barriers, extra-absorbing barriers for mopping up resulting spills.

£99

Epsilon Magnifier

Sick of struggling with tiny epsilons? The revolutionary new epsilon magnifier simplifies analysis by increasing all epsilons to values > 1.

£δ

Set of Pathological Cases

For the more experienced traveller, save money with our nowhere-dense set of luggage.

£40

Quantum Surfboard

Don't use on unrestricted wavefunctions

£35

Calculus-removing toothpaste

Guaranteed opaque.

£5

Anthropomorphiser

Ascribes human qualities and emotions to functions, sets, numbers etc. Not for use on mathematicians!

£*i*

The Escher Machine

The ball rolling down an infinite slope generates enough energy to power a light bulb.

NEW! Uphill version: uses two 1.5V AA batteries per day. The ideal gift for someone you dislike.

£ ∞

All items in our catalogue can be ordered by writing to

The Archimedeans
Ω Kolmogorov Street
X1024 Cantortown

Invented by Chris Cummins, Eureka 56

CALL MY BLUFF

Compiled by C J Budd, Eureka 43

Here you see three different definitions for some obscure mathematical terms.
It is your task to find – with justification – the correct one. Solutions are on page 94.

The Tarry Point

(1) This point was discovered by that well-known mathematician Nicolas Tarry. Given an earth-moon-sun-spaceship system, the Tarry point is that point where a body would be in equilibrium.

(2) The Tarry point of a triangle is the point on its circumcircle opposite to its *Steiner point* – the point of intersection of the lines through vertices of the triangle parallel to the corresponding sides of the first Brocard Triangle. The vertices of the *Brocard triangle* are on the points of intersection of the lines from the vertices of the triangle to the *Brocard points* X and Y. These are such that $\angle XAV = \angle XAC$ and $\angle YBA = \angle YAC$.

(3) Given a dynamical system, the Tarry point is the point at which the rate of growth ceases to be exponential – although polynomial growth is still permitted.

A Room Design

(1) A system of organising movements in a Bridge tournament, formalised by Mr Room.

(2) A dissection of a square into smaller squares of different side lengths ('rooms'), discovered by Trinity students in 1939.

(3) Tile a cuboid in $\mathbb{R}^n$ regularly by sub-cuboids. Mark certain faces, ensuring that no subcuboid has >3 faces marked. Then if it is possible to go from one subcuboid to another entirely by marked faces, we have a room design, where the subcuboids are the rooms and the marked faces are the doors.

Unger's Translation

(1) A device discovered by Unger and widely used in the engineering industry. It transforms a problem in potential theory to another which may be easier to solve.

(2) Ungers Translation can transform a series of simultaneous nonlinear partial differential equations to non-Euclidean geometry, where it looks prettier even if it still may be insoluble.

(3) Ungers' translation is, of course, a translation by Unger of a work by Hilbert.

A Mouse

(1) On each face of a tetrahedron construct another tetrahedron of side 1/3 of the original. Continue this process for ever. What you end up with is a mouse: a finitely small yet infinitely furry little animal.

(2) A mouse is, naturally, a subset of a Cat, a connected absorbing topology! A mouse is any subset of a cat which has a tail (i.e. a proper one dimensional subset. This tail must of course be unique and no two mice are permitted to have the same tail.

(3) A premouse is an admissable set with an ultrafilter which thinks the ordering it gets from the ultrafilter is a well-ordering. If the ordering is close enough to let us iterate on the ultrafilter we have an iterable premouse. If this is well behaved we have a critical iterable premouse. A mouse is a critical iterable premouse for which every sub-premouse is also critical.

Solutions

Archimedeans Problems Drive

1 Dazzling Dice

Most likely is 6, with probability 16807/46656.

2 Snappy Surds

20, 28 and 100.

3 Painful Primes

999 917 is prime.

4 Compelling Convergence

(a) diverges;
(b) converges to $\pi^2/6$;
(c) converges to $\tanh^{-1}(\log 2) - \log 2$.

5 Superb Sets

(a) is countably infinite;
(b) is finite of size 1;
(c) is uncountable as it bijects with $\mathbb{R}$;
(d) is uncountable as it bijects with $\mathbb{R}$.

6 Triumphant Treasures

The treasure is buried on the moon.

7 Curious Coins

You want to go first for all n.

8 Perceptive Polygons

Yes!

9 Terrible Triangles

$\left(\frac{8}{19}, \frac{2\sqrt{3}}{19}\right)$.

10 Rough Relations

$$f(n) = \begin{cases} 0 & n \le 3 \\ 2 & n = 4 \\ \frac{1}{2}n(n+1) & n \ge 5 \end{cases}$$

11 Gorgeous Geometry

Line perpendicular to OD.

12 Mysterious Matchings

2

13 Dazzling Digits

4000

14 Cryptic Crossword

The treasure is buried on the moon.

Call My Bluff

The Tarry Point: Definition 2
A Room Design: Definition 1
Unger's Translation: Definition 3
A Mouse: Definition 3

Copyright Notices

Articles

Page 6: Talking to Computers © Stephen Wolfram, published as *'Programming with Natural Language Is Actually Going to Work'* on his blog, November 2010

All other articles are © The Archimedeans.

Illustrations

Cover Design
Andrew Ostrovsky

Inside Front Cover: Eights Sculpture
© George Hart, *georgehart.com*

Page 4: Top Right and Bottom Left Photos
© Shubnit Bhumbra

Page 6: Computer Brain
© Depositphotos / agsandrew

Page 10: Fortress and Sun
Paul Klee, *Public Domain*

Page 21: Teacups
© Depositphotos / DepositNovic

Page 24–27: Backgrounds
© Depositphotos / lolaferari

Page 29: House Numbers
© Depositphotos / dip2000

Pages 30–31: Surface
MorgueFile User Paul Schubert

Pages 32–37: Mandelbrot Images
© Philipp Legner

Page 32: Fractal Fern
Wikipedia User 'Olegivvit', (CC BY 2.5)

Pages 38–41: Statistics Charts
© Depositphotos / Khakimullin

Pages 42–43: Playing Cards
© Depositphotos / galdzer

Pages 44–45: Sound Waves
Created by Philipp Legner, based on graphics © Depositphotos / jineekeo

Pages 47, 77, 81 and 85: Photos of David Hilbert, Frank Ramsey, G H Hardy and Paul Erdos
Public Domain

Pages 48–49: Background
NASA / ESA / Spacetelescope.org

Pages 48–49: Photos
Public Domain, CERN, Individual Mathematicians

Page 50: Plasma Ball
© Depositphotos / crazycolors

Pages 54, 56–57: Space Images
NASA, ESA, S Beckwith, Hubble Heritage Team

Pages 58–59: Multiverses
© Depositphotos / maninblack

Page 62: CM Background
NASA, ESA

Page 64: Particles
© Depositphotos / prill

Pages 68–70: Photos
© Berman and Davey

Pages 72–73: Glacier
Daniel Schwen (CC BY-SA 2.5)

Pages 74–75: Glacier
Wikipedia User 'S23678', (CC BY-SA 2.5)

Pages 72–75: Research Photos
© Indranil Banik and Justas Dauparas

Page 82: Nuclear Bomb
Public Domain

Pages 83: Mitered Fractal Tree
© Foundation MathArt Koos Verhoeff
Used with permission

Page 87: The Ultimate Painting
Drop Artists, *Public Domain*

Pages 88–89: The School of Athens
Raphael, *Pubic Domain*

Pages 89: Book Cover
© Cambridge University Press

Pages 90–91: Book Covers
© The respective publishers

Page 92: Images
stock.xchng users tunnelvis, hisks, garytamin, Cieleke, purdywurdy and stuarrose; MorgueFile user pschubert; Wikimedia user Lethe (CC BY-SA 3.0)

All other images and diagrams are © The Archimedeans.

If you would like to reproduce any articles of graphics in this publication, please email archim-eureka-editor@srcf.ucam.org

Eureka for iPad

Coming to the App Store in December.

Join The Archimedeans

Join one of the oldest student societies and get free entrance to countless amazing talks, great social events, discounts in our bookshop and three free copies of Eureka!

Membership is only £5 per year or £10 for life.

Email *archim-secretary@srcf.ucam.org* for details, visit *www.archim.org.uk* or write to

The Archimedeans
Centre for Mathematical Sciences
Wilberforce Road
Cambridge, CB3 0WA
United Kingdom

Write for Eureka

If you want to contribute to future issues of Eureka, or be on the editorial team, please email *archim-eureka-editor@srcf.ucam.org*. Further details can be found on our website.

Subscribe to Eureka

Send us a check along with your own postal address, or subscribe online at *www.archim.org/eureka_subs.php*.

For more details, or to order previous issues of Eureka, please email *archim-subscriptions-manager@srcf.ucam.org*.


EUREKA
EUREKA
EU RE KA
eureka
eureka
EUREKA
eureka
EUREKA
EUREKA
EUREKA
EUREKA
EUREKA
EUREKA
EUREKA
EUREKA 47
EUREKA
EUREKA
EUREKA
eureka 51
EUREKA
EUREKA
Eureka 55
Eureka
EUREKA
EUREKA
eureka
eureka