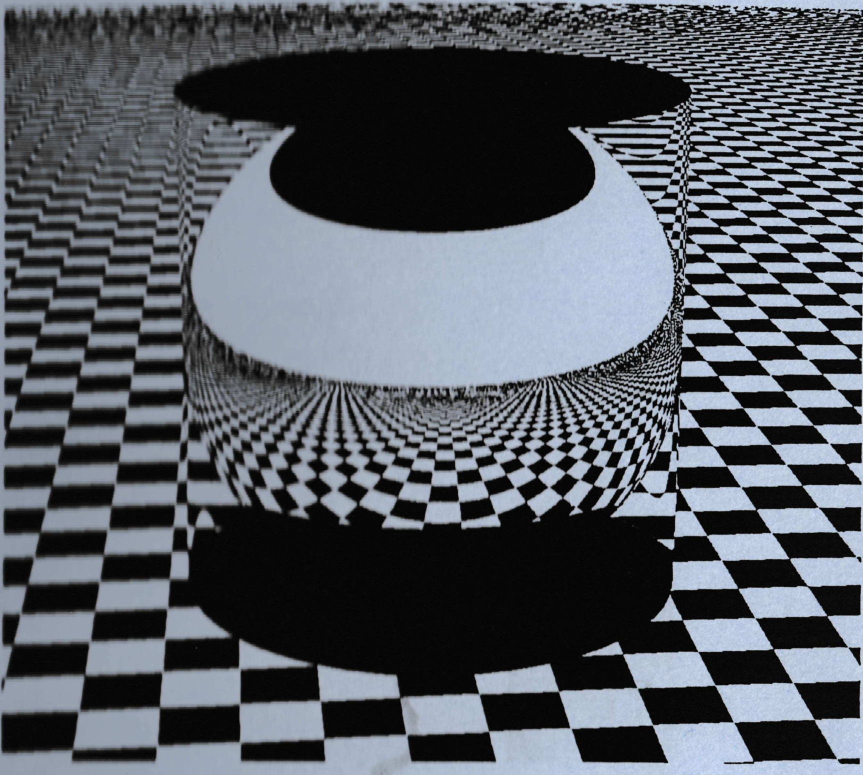
EUREKA



The Journal of The Archimedeans

Number 50 April 1990

Eureka

Eureka is the journal of the Archimedeans, which is the Cambridge University Mathematical Society, and a Junior Branch of the Mathematical Association. *Eureka* is published approximately annually, but since it, like the Society, is run entirely by student volunteers, it is impossible to guarantee precise publication dates. The Society also publishes QARCH, a problems journal.

Subscriptions

A subscription account may be opened by sending at least five pounds to the Business Manager at the address below; issues will be sent as they are published and you will be informed when your account is running low. It is also possible to subscribe to *Eureka* by standing order. Back numbers of some of the more recent issues are available, and photocopies of any issue may be made available by special arrangement. TEX source for issues since issue 49 may also be available, subject to copyright; contact the Business Manager for details.

Electronic Mail

The Society can receive electronic mail via JANET, ARPAnet and so on, and can read most electronic media. Our electronic mail addresses are given below. There may be long delays in dealing with messages of any kind in the University vacations (especially June to October), and you are asked not to send large amounts of text electronically without prior warning.

Addresses

Correspondence, clearly marked "Editor" or "Business Manager", should be sent to:
Eureka, The Arts School, Bene't Street, Cambridge CB2 3PY, England

Electronic mail from the UK should be sent to:

`archim@uk.ac.cambridge.phoenix`

From the rest of the World, the following should work:

`archim@phoenix.cambridge.ac.uk`

But if it doesn't, try whichever of these seems the more appropriate:

`archim%phx.cam.ac.uk@nss.cs.ucl.ac.uk`
`archim%uk.ac.cam.phx@ukacrl`

EUREKA

Number 50 — Editor: Mark Wainwright — April 1990

Table of Contents

Acknowledgements		2
Editorial		3
The Archimedeans	*Colin Bell*	7
Tris	*Adam Chalcraft*	9
Bulgarian Solitaire	*Thomas Bending*	12
The Creative Archimedean At Work		20
Call My Bluff	*Stephen Turner*	21
An Important New Result		23
Perestroika in the Maths. Faculty	*Heather Mendick*	24
Problems Drive	*Frazer Jarvis and Graham Nelson*	28
Beyond the Vanishing Point	*Ian Stewart*	34
What Happens when an Engineer Learns some Maths	*Piers Bursill-Hall*	41
G.H. Hardy and Control Engineering	*Jonathan R. Partington*	57
Sir Michael Cometh		59
Unfair Games	*Martin Baxter*	60
A Logical Puzzle	*Colin Bell*	68
"Debased" Crossnumber	*Quinapalus*	69
On Large Numbers	*Mark Wainwright*	70
The Wobbler	*Frederick Flowerday and David Singmaster*	74
Clerihews	*G. S. Murchiston*	78
The Professor's New Book	*Kevin Murphy*	79
Book Reviews	*John Aspden, Tim Auckland, Mark Wainwright*	87
More Call My Bluff	*Stephen Turner*	90
Solutions to Problems Drive etc.		91
The Cover	*Thomas Bending*	Inside back cover

ISSN 0071-2248

Archimedeans' Committee 1989-1990

ACKNOWLEDGEMENTS

It is a pleasant duty to record the vast amount of help received during the exhilarating work of editing *Eureka*. The foremost debt is as usual to the contributors; I should add that they have without exception been more helpful than an editor has any right to expect in the preparation of their articles for final publication. The journal was typeset almost entirely in TEX, and in this regard incalculable thanks are due to Adam Chalcraft, Robert Hunt, George Russell, and too many others to name. Those diagrams too horrible to consider in TEX were executed in CAMPLOT, and on this and many other fronts I am deeply indebted to Colin Bell.

I am especially grateful to Dr Mark Manning of the Department of Applied Mathematics and Theoretical Physics for considerable help with the production of laser-printed hard copy, including use of the DAMTP laser printer; and to Graham Nelson, editor of *Eureka* **49** and Patriarch of the Society, for his unfailingly patient and helpful advice and paternal manner on a number of occasions. Thanks too to Dr Martin Barlow for casting a brief eye over the article "On Large Numbers" to check that it was not pure drivel; any remaining errors are naturally entirely mine.

Finally, to all those supervisors for whom, during the Lent term, the Editor did (even) less work than is his wont, a vote of heartfelt gratitude for their uniform patience.

Editorial

The editors of the first issue of *Eureka*, published in 1939, stated clearly why they considered it a useful venture. Student mathematicians were hitherto, they argued, "without a medium for stating their views, for discussing their present training and their future prospects, for publishing their less orthodox researches. They had no encouragement to learn of new discoveries and new publications, to see, in historical perspective, the great developments of days gone by." And despite a modest disclaimer, their issue fulfils these ideals admirably well. *Inter alia* they print a long extract from a letter received from the University of Chicago in support of the new publication. "There is surely a need," this remarks, "for a publication addressing itself to the less technical and more cultural aspects of mathematics."

If this description of *Eureka* is a foreign one to the ears of more modern readers, the fault is not with its editors. Eddy Welbourne, for example, in his editorial in *Eureka* **46** (1986) remarks

> I would hope that at least some of those with interests other than straight mathematics will pluck up the courage to write for *Eureka* ...
>
> If they do not, we will all continue to lose out.

and the same sentiment is reiterated by a number of recent editors.

The difficulty of attracting articles which are not primarily mathematical continues. Nor is it a new one. Terry (now Professor) Wall, in his editorial in *Eureka* **20**, charts the rise of the problem. He suggests that the effect of involved mathematical articles (which appeared from about the 8th issue) was "to overawe those with little confidence in their mathematical ability and inhibit them from writing anything at all for EUREKA." He goes on:

> But there is a need for discussing the broader functions of maths. in society even nowadays. ...
>
> Our contact as a body of mathematicians with other faculties or with other universities or even with members of our own faculty board is practically non-existent ... In fact we have not really bothered to form ourselves as a body of mathematicians, and so need not expect to have any influence over matters of concern to us. But nobody seems to be concerned so perhaps it doesn't matter.

Certainly the *Eurekas* of the late 1970's furnish little evidence that anyone was concerned. They highlight too another difficulty, in that there is a severe shortage of articles written by students at that time; issue 40 is an extreme example of this.

Articles by fellows are of course a feature of *Eureka*, and have been since the third issue, where the editors wrote "we have in this issue for the first, but not, we hope, the last time, articles written by Senior Members of the faculty." But in issue 6 the editors note that they feel almost like writing, "We have in this issue, apparently for the last time, articles written by Junior Members of the University," and the problem has been endemic ever since. "To speak bluntly, the Archimedeans cannot expect to have a Journal if they do not write for it," remarks one D. Monk,† speaking

† Whose name will be familiar to at least some Archimedeans.

bluntly in Issue 19. He continues "A feature of the present issue is the complete lack of poetical contributions, so we would remind you that, whilst we appreciate articles of serious mathematical content, contributions in a lighter vein are also acceptable." And according to G.C. Shephard in issue 11, "Articles for *Eureka* are still urgently required—particularly those of a trivial or humorous variety."

But it is not enought to quote past authorities in support of such opinions. Times change, after all, and no doubt *Eureka* should change with them. Perhaps the notion of a journal addressing itself to the cultural and humourous aspects of mathematics, with articles mainly by undergraduates, is an outmoded one, one for which there is no longer a call. The President remarked in issue 1 that we should "make it interesting to every Cambridge mathematician"; but that was half a century ago, and perhaps today we shouldn't do anything of the kind.

Yet I conjecture with modest confidence that few readers of *Eureka* would see it transmogrify into a mere mathematical journal. For one thing, there already exist a superabundant sufficiency of such publications. Why should the Archimedeans burden itself with the (very considerable) expense of a journal which had been better published by the Department, or by the University Press, or by the Cambridge Philosophical Society, or indeed not at all? If the aim is merely to furnish such a journal to members of the Society, why not distrubute the Journal of Recreational Mathematics to its members, bought in bulk at a discount from the publishers? That these are rather silly questions is immediately obvious.

What function, then, should *Eureka* fulfil? Primarily, of course, its purpose is to entertain and inform the society's members. Now as it stands this is an empty remark, for it tells us nothing about what the journal should contain. Let us see if we can extract some content from it. Firstly, then, let us see about entertaining the members.

Mathematicians have a sense of humour.† They have too a noted affinity for nonsense-verse.‡ Hence humorous articles in general, and nonsense-verse in particular, are appropriate for inclusion.

Students at University are generally there by choice, and they have chosen to be there because of an interest in their subject. Maths. faculties provide an education in the subject itself, but it remains up to the students to put this into a historical and a social perspective. Here is another area where *Eureka* can usefully entertain and inform.

I have yet to meet an Archimedean without diverse non-mathematical interests and strong opinions on something. *Eureka* serves as a vehicle for those with views or mathematical ideas to express, allowing them to share the former and discuss the latter.

However; *Eureka* is not only a method of communication between Archimedeans. It is a costly and a time-consuming alternative to talking to one another, a pastime to which members of the Society have every opportunity and frequent recourse to indulge. For communication on a broader scale, too, the Archimedeans have an invaluable institution in the form of lunchtime meetings, where informal talks are given (generally) by undergraduates; though these have all but died out in recent years, much to the detriment of the Society. Let us hope to see more of them in the year to come.

---

† If often a somewhat eccentric one. This, of course, is only another reason why it should be catered for specifically by a journal such as *Eureka*.

‡ It was noted when the results of the Archimedeans' Questionnaire were printed (issue 2), and again when the questionnaire was reprinted (issue 14; results in issue 15).

From entertaining we pass to informing the members of the Society, that is, serving as a link between the Archimedeans and its members on the one hand, and on the other, in some sense, the "outside world". Let us find a few examples of what constitutes the outside world.

The maths. lectures and Tripos are there for our benefit. Matters concerning the Tripos are therefore naturally of interest to the Archimedeans who should, therefore, have some means of expressing their opinions and keeping abreast of matters. In this issue Heather Mendick, the former Faculty Rep., has written about the forthcoming changes to the Tripos necessitated by changes in secondary education.

Then again, if Archimedeans are to make informed decisions about their course they should have a basis for comparison and contrast; so "the outside world" includes, too, maths. faculties elsewhere.

Again, most Archimedeans have been at school within the last three years, and some will end up teaching maths. Thus the world of education must be added to the list. It should be noted that all of these remarks work both ways. As we will be interested in the manœuvres of the Faculty Board, so they will, of course, want to gauge the climate of student opinion from time to time. Similarly we may assume that students elsewhere will have some interest in the Cambridge Maths. Tripos. This all epitomises *Eureka*'s final and crucial rôle, which is as the public face of the Archimedeans—the front of the shop, so to speak.

We live in a time when, increasingly, pure research—of whatever kind—must justify its existence or suffer severe loss of resources and, at worst, disappear; this state of affairs is at once ludicrous and alarming. The philosophy is apparently as follows: the academical prowess of the nation is a means by which to achieve its fiscal security. Thus, research—and indeed education—should be funded precisely in accordance with its foreseeable industrial, or other financial, return. The logical extreme of this approach would presuably be highly selective funding of the sciences, and almost none of the humanities.

But, quite apart from the fact that there is probably no correlation whatever between a project's foreseeable return, and its actual long-term benefit; such a programme fails because in the scramble to fulfil the means it quite loses sight of the ends. A nation's government seeks its prosperity not, presumably, as an end *per se*, but rather as a means to the happiness of its subjects. But prosperity, while it can make us tolerably sure of being comfortable, in no wise furnishes any guarantee of happiness. Aristotle observed that "all men possess by nature a craving for knowledge", and mere material comfort is not sufficient for happiness when this intellectual curiosity is uncatered for. Housman remarks,

> Let a man acquire knowledge not for this and that external and incidental good which may chance to result from it, but for itself; not because it is useful or ornamental, but because it is knowledge, and therefore good for man to acquire.

At any rate; the right to an education should rank with the right to political freedom as one of our most fundamental and most valued prerogatives. Everyone should have the chance of a full education, irrespective of ability to pay for the tuition, and of whether the chosen course of study is one which (like mathematics) widens one's career prospects, or one which is of purely intellectual interest and makes no difference to one's likelihood of obtaining employment or one's ultimate salary.

However; the reality is unfortunate but quite different. Of course nothing the Archimedeans might do will have the slightest influence over the well-being of the maths. departments here, which in all honesty does not at the moment look greatly threatened, or indeed elsewhere. What it may do is to give some idea to non-mathematicians of what mathematicians in general, and the Archimedeans in particular, do, and what motivates them.

This is not the place to enter into the motivating elements of mathematicians in general; and I hope I have made my views on those of the Archimedeans sufficiently clear. If I seem to have burdened the shoulders of *Eureka* with all the troubles of the world, I freely confess that I have set an impossibly high ideal. It is not intended, however, as a genuine standard by which to judge the journal's success, but rather as a pattern which should influence the nature of articles sought for inclusion.

It may have occurred to some readers that the activities of the Archimedeans are in general rather less broad than those of the Society I have been describing. To see why this might be, let me return to the Objects of the Society quoted by the president in *Eureka* **3** (1940): our objects were set forth in the rules as

> to study the social, cultural and economic aspects of mathematics and its relation to the other sciences, and the history, philosophy and teaching of mathematics in various countries.

I set it forth here that it might be contrasted with its more pedestrian counterpart in the current Constitution. There Chapter II ("Objects") reads

> SECTION 1
>
> The Objects of the Society shall be:
>
> (i) to promote the study of Mathematics and to further the cause of Mathematicians in the University and elsewhere;
>
> (ii) to cooperate with, and to cause cooperation between, the College Societies;
>
> (iii) to cooperate with the Faculty of Mathematics in the University;
>
> (iv) to hold lectures and other events of interest to Mathematicians;
>
> (v) to publish two journals, which shall be called "EUREKA" and "QARCH", and other publications of mathematical interest;
>
> (vi) to operate a Bookshop for the benefit of its Members.
>
> SECTION 2
>
> The activities of the Society shall have no political, religious or racial bias.

Certainly these current objects are far narrower than those earlier ones. I seem to hear the objection, "but it is not a society's written constitution, but rather those who constitute it, that affects the spirit is which it is run." Certainly this is true to some extent. On the other hand the Archimedeans enjoys remarkable degree of continuity in despite of an almost complete change in the active membership every three years. The difference between a more and a less active committee will be seen as due to the one being particularly energetic, or the other being particularly apathetic, depending on the aims which the Society sets itself in the Constitution. It is an enthusiastic committee and membership coupled with constitutional change which will bring about a more lasting broadening of the base of the Society's activities.

The Archimedeans

Colin Bell (Chronicler 1989–90)

"The Archimedeans have had another successful year ..." seems to be a favourite thing to say among Chroniclers sitting down with a blank piece of paper thinking of things to say about the previous year. I see no reason not to say the standard thing—after all, it has the dubious merit of truth. The year 1989–90 may not go down as a great vintage in the history books, but it was a fairly good one nevertheless. The committee did whatever it is that committees are supposed to do, and did it reasonably well too. So let us cast our minds back to that ridiculously hot summer of 1989 ...

In the Easter term, it is well known that nothing matters until the second week of June, and 1989 was no exception. The first event in the term of note was the Garden Party, which saw temperatures in the 90s and a correspondingly high turnout. The Barbershop Subgroup sang sweetly (although not sweetly enough for one senior member of the College, who was observed shutting his window when the singing began, and opening it again after it finished) but were thwarted from their usual method of arrival (a punt) by the Landscape Gardening Subgroup's inability to divert the Cam through Pembroke. The Garden Party was followed in quick succession by the Punt Trip to Grantchester, the ramble, including a scenic crossing of the M11 (at road level), the Croquet Match against DPMMS (which, with two of their players being on international duty and a blossoming of Peterhouse talent, we managed to win for once) and a Punt Tiddleywinks match against the Dampers, who invented the sport. This we managed to lose.

The summer over, we returned to Cambridge; four months older and perhaps slightly wiser. The Societies' Fair (with the now traditional computers) and the Squash brought in a fairly average number of freshers, eager to start their first term of university mathematics (one wonders how keen they will be three years later). The term proper began with one of the best attended meetings in recent years, given by Professor Penrose of Oxford, linking together such varied fields as logic, neurology, quantum mechanics and cosmology, in a manner loosely connected with his new book (of which a review appears elsewhere in this issue). Further talks were given by Professors Olive, Ledermann, Donaldson and Dalitz and Drs Monk, Berkshire and Isenberg. The last of these was a multivisual experience; the centrepiece of the show being a large plastic dustbin filled with soapy water, from which he proceeded to blow vast numbers of incredibly-shaped bubbles (or credibly-shaped bubbles, depending on how good you are at geometry). The only pity was that more people didn't come to it, a problem which seems to be endemic amongst the Maths. Societies at the minute.

However, the highlight, at least, of the Michaelmas term, was the Call My Bluff competition. Teams from Imperial, Nottingham, Bristol and King's College London arrived in Cambridge, and the first three did battle with the native Cantabrigians in the splendor of Trinity OCR. King's went sightseeing. As a result of subtle changes in the scoring mechanism from the Chair, Cambridge ended up with no points at all. Those wishing to know further details are pointed in the direction of the article in this issue by Stephen Turner.

And so the Society reached the Annual Mark Owen Expulsion Meeting, at which the membership expel Mark Owen from the Society and then reinstate him. This done, the meeting pauses only briefly while the old committee is sprayed with liberal quantities of white foam, before it sobers slightly and elects the new committee. That elected for 1990–1 has a refreshing reduction in the number of members from Trinity (from six to four—it seems the Society had an allergic reaction to the comment J. Rickard made in *Eureka* **43**, "...for the first time in the memory of even the oldest undergraduate, there were no members of Trinity College on the new committee!" and it has elected at least four at every AGM since). It does however have the unfortunate drawback of including only two first-years; it would be nice to see more of them about, since it is they who will be running the Society from next year.

As for the College Societies, they appear to be going well despite very poor attendances at some meetings. The New Pythagoreans managed a meeting at which the audience consisted solely of the current and two previous Secretaries and one "friend" of the Society. (I was the friend, and it was in fact a very good talk.)

The subgroup situation is mixed. The Barbershop Subgroup plan to make vaguely musical noises again this Summer; the Othello Subgroup, which has bifurcated into two rooms this year, is thriving and the Puzzles and Games Ring likewise. The PGR has also split into two rooms, but one of them is that of Il Dottori Adam Atkinson of La Sapienza University in Rome. The PGR's most insoluble puzzle of the year has been to explain why vast hordes of people arrived just before a Plenum meeting held in the same room and went afterwards, rather than the other way around. The Musical Appreciation Subgroup has appreciated somewhat less music this year; Philip Belben has retired after many years' service and his successor Mark Allan had to wait a term in the outer darkness of Trinity lodgings before getting a room in the new Blue Boar Court. Much is promised from this new location, however. The Mathematical Models Subgroup has suffered similarly since its proprietor, the multi-faceted Tim Auckland, was moved out into the wastes beyond Newnham.

Thanks are due to the four speakers who spoke too late to make last year's column—Dr Hawkes, Mr Hersee, Professor MacCallum and Dr P. Johnson-Laird, and also to our ex-Senior Treasurer, Dr Jonathan Partington, who retired after two years in the post last summer because of his impending defection to Leeds. He has been replaced by Dr Imre Leader of Peterhouse—long may he reign.

So what is lined up for next year? Not a great deal as yet, but plans are being drawn up for the 1991 Triennial Dinner which is likely to be held in February of that year. As usual it would be nice to have some past members present— anyone interested is invited to write to the Secretary.

It only remains for me to say thank you to the other seven committee members of last year for making it fun, and to wish the new committee good luck in their endeavours to make 1990–91 "another successful year".

Tris

Adam Chalcraft

NOTE. There is a glossary of terms at the end of this article.

What is Tris? Tris is to Tetris† what 3 is to 4. If you don't know what Tetris is, firstly congratulations are in order—how did you manage it?—and secondly you had better read the next paragraph to find out. If you already know Tetris, you can skip the next paragraph.

Tetris is a computer game—player *vs.* computer. It is played on a big (10 across by 20 high)‡ Connect-4 board, with totally different pieces and rules. Tetrominoes (which are to dominoes what 4 is to 2) come in at the top of the board and slowly drift down the board until they hit something, when they stop. The player can move each piece left or right as it drifts down, or rotate it 90° anti-clockwise. Tetris has 7 different pieces. See the glossary. Once a piece has stopped, a new randomly chosen piece comes in from the top. If ever there is a complete row of filled squares, it disappears and everything above drops one square. The idea is to keep going for as long as possible. You ought to be warned that Tetris is mildly addictive in the same way that nuclear missiles are mildly dangerous.

I was interested in the question as to who had the win. Either the player can keep going for ever, no matter what pieces fall, or else the Tetris daemon, playing Devil's advocate, can choose the pieces (dependent on what the player does) so that the player loses.

After extensive research, conducted in the PGR research institute, it was eventually conclusively proved that the answer to the question "Is Tetris a win for the player?" is "Probably not". Undaunted by this minor set-back, I journeyed into the frozen North to further contemplate this deep question, and I can now proudly present the following.

THEOREM. *Tris, played on a board $n \geq 2$ wide by $m \geq 5n - 3$ high is a win for the player.*

For the proof, I need a board notation, and I am going to use a rather strange one. It assumes the position is grommet-free, and the heights of the stacks are monotone (not necessarily strictly) increasing left to right. The position is then described by the lengths of the *rises* and *treads* of the resulting staircase, read from the bottom left corner to the top right. The sequence must start with a tread, otherwise there would be a complete row. It will end with a rise, which is the right-hand wall.

I claim that a position can be maintained at all times in which all the rises are at most 5, except for the first one, which is at most 6, and the last one, which could be as large as m. Moreover, if the first rise is 6, then the first two treads must both be 1. A position satisfying this lot is called *green*.

This will always fit into the $n \times m$ board, with an empty row at the top, and the starting position (tread n rise m) is monotone and green, so this will prove the theorem.

† Tetris is a Trade-Mark.

‡ This information comes from the Macintosh DA version. Available at a user area near you.

There are 2 pieces in Tris, and .

So we are in such a position, and in falls a , say. Since $m > 2n$, there must be a rise of at least 3 somewhere. Put the , rotated, next to the leftmost rise of height at least 3. The resulting position is clearly monotone, the move can create a rise of height at most 5, and if the first rise was 6, then the move makes it 3, so the position is still green.

Now we have to cope with a . We will put this in either as , or as .

Abbreviate rise to r, and tread to t, and let 2+ stand for any number greater than 1. I claim that the position satisfies one of the following:

(i) It contains (t2+ r2+),

(ii) It contains (r1 t1),

(iii) It starts with (t1 r? t1).

For suppose not. Then from (iii), the sequence has (t2+) as either the first or second tread. It must continue (t2+ r1 t2+ r1...) to avoid (i) and (ii), but then the sum of the rises is less than m, which is a contradiction.

So one of (i)-(iii) holds, and we have to put the somewhere. We find the leftmost of (i)-(iii) which occurs, and put the piece in there. In a (t2+ r2+), we lay it in as . In a (r1 t1), we fit it on the step as . Finally, in the case where the sequence starts with (t1 r? t1), we put it in as , removing 2 rows. That move was TCMS if the sequence started (t1 r2+ t1).

I now only have to show that these moves (i)-(iii) preserve the monotone and green nature of the position. Monotonicity is obvious. If there is no rise greater than 5, look back 2 paragraphs, and note that except for the start (t1 r2+ t2+), we find an occurrence of one of (i)-(iii) before or including the first (r2+), and so the move can only create a rise of height at most 3.

For the start (t1 r2+ t2+), there cannot possibly be a problem except in the case (t1 r5 t2 r2+), but this becomes (t1 r6 t1 r1), which is still green. Thus the proof is complete. $\square$

Note: Careful study of the proof reveals that every move was effeminate, and every move which took out a row was Richards.

OPEN QUESTIONS

1. The Big Question. Who has the win in Tetris?
2. Can you win at Tris without using TCMS moves? Note: If you can do, say, $n = 10$ and $n = 11$ without TCMS, then you can do all $n \geq 90$.
3. Suppose, in Tetris, the daemon has to announce what the sequence of pieces he is going to drop will be at the start of the game. Who has the win now?
4. Can I have my free copy of *Eureka* now, please?

My personal guesses are: The daemon, No, The player, and Yes.†

† The author's conjectured answer to question 4 has since been confirmed by a simple constructive proof. –Ed.

GLOSSARY

Disclaimer — This glossary is probably inaccurate, and definitely incomplete, since new terminology is constantly being invented to cope with new situations.

de-grommet *vti.* To remove (a grommet) by skilful play, or otherwise.
effeminate *adj.* (of a piece) Taking up as many rows as possible. Also (of a play) involving the dropping an e~ piece. cf. **virile.**
grommet *n.* A hole in the position. A connected set of squares not connected to the outside world.
I *n.*  (in Tris);  (in Tetris).
L *n.*  (in Tris);  (in Tetris). Hence L′ .
O *n.* .
outside world *n.* The squares in the top row where new pieces come in.
perfect *adj.* (of a move) Entirely destroying the dropped piece, as a result of complete rows being formed.
pre-TCMS (decline as **TCMS**) Grommet.
red *n.* I. Hence **r~-dependency** *n.* A need of r~s to get out of trouble. A common situation for beginners to be in.
Richards *adj.* (of a move) Removing as many rows as possible with the dropped piece. As in "Richards is perfectly effeminate".
S *n.* . Sometimes called Z′.
T *n.* .
TCMS *adj.* A **T~ move** One in which a grommet is made which immediately disappears because a complete row is formed. Also *n.* A T~ move. *vi.* To make a T~. *vt.* (of a piece) To drop in a T~ move.
virile *adj.* (of a piece) Taking up as few rows as possible. (of a move) involving the dropping a v~ piece. cf. **effeminate.**
Winston *vt.* (of a piece) To slide under another piece by moving left or right at the last moment.
Z *n.* . Sometimes called S′.

Bulgarian Solitaire

Thomas Bending

Consider the following game:
Deal a deck of n cards into an arbitrary number of piles, with an arbitrary number of cards in each pile. Take the top card from each pile and put the removed cards together to form a new pile. Repeat this move indefinitely.

For some reason this exciting pastime is known as *Bulgarian Solitaire.* We can ask a number of straightforward questions about it:

1. Since the number of positions that the game can reach is finite (it is the number of partitions of n) the game must eventually reach a position which has occurred before, and so enter a loop. How many such loops are there, and can we characterise the positions in them?
2. In particular, if $T_k = \frac{1}{2}k(k+1)$ denotes the k^{th} triangular number, and $n = T_k$, there is an obvious one-step loop whose only position has piles of $1, 2, \ldots, k$ cards—call this position Δ_k. Is this the only loop for this n?
3. If we define the *life* of a starting position to be the number of moves which must be performed in order to reach a loop, is there a good bound on the maximum possible life of a position?

It turns out that for $n = T_k$, the one-step loop is the only loop, and thus all starting positions eventually lead to this state (proved by Jørgen Brandt in 1981 (see [1])). This is a special case of the characterisation of possible loops for general n given below. The same article mentions the conjecture that, for $n = T_k$, the maximum possible life of a pattern is $k(k-1)$, and that this has been confirmed by computer up to $k = 10$ (by students of Donald Knuth). We will show that this bound is the best possible, i.e. that there is always a starting position with this life.

Two convenient ways of representing a general position are as follows:

- As an ascending sequence of numbers denoting the heights of the piles, e.g. $[1, 2, 3, 4]$ for Δ_4. We can write $[1, 1, 1, 2, 3, 3]$ as $[1^3, 2, 3^2]$ for brevity.
- As left-justified rows of blobs, with one row per pile and one blob per card, on the vertices of a doubly semi-infinite integral lattice, so that $[1, 2, 5]$ is represented by

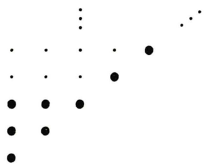
$$\begin{array}{cccccc} & & \vdots & & & \\ \bullet & \cdot & \cdot & \cdot & \cdot & \\ \bullet & \bullet & \cdot & \cdot & \cdot & \cdots \\ \bullet & \bullet & \bullet & \bullet & \bullet & \end{array}$$

We say that such a diagram is *legal* if no row contains a blob to the right of a space, and *neat* if the rows are in ascending order of length from top to bottom, i.e. if no column contains a blob above a space. We shall only ever consider legal diagrams. We shall sometimes use North to mean up, East to mean right, and so on.

Given a neat diagram, a move in the game consists of removing the left-hand column of blobs, rotating it to form a row, placing it below the bottom row of the diagram so that its left-hand end is below that of the other rows, and permuting the rows if necessary to form a neat diagram (we must restore neatness, lest later moves lead to an illegal diagram).

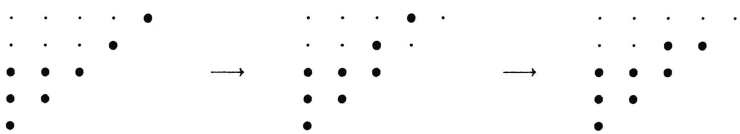
If the new diagrams are placed so that their bottom left blobs coincide with that of the original diagram, then (after a few minutes pushing coins around) it is clear that in the first part of this move all the blobs in the original diagram except those in the leftmost column have been translated one place NW, while those in the leftmost column have been translated as far as possible SE:

To make this movement easier to follow, we shear the diagram so that the movement is horizontal, i.e. fix the leftmost column, move the next column to the right up one place, the next column up two places and so on, so that $[1, 2, 5]$ is now represented by



Note that in this format a diagram is legal if no SW–NE diagonal has a blob to the NE of a space, and as before a diagram is neat if no column contains a blob above a space.

A move now consists of moving each blob one place to the left in its horizontal row, those in the leftmost column going to the right-hand end, and then permuting the SW–NE diagonals to restore neatness, e.g.



It is convenient to number the rows $1, 2, \ldots$ from bottom to top, so that row r contains r blobs and spaces, but to number the places within the r^{th} row $0, 1, \ldots, r-1$, so that the position of a blob in row r is reduced by 1 (mod r) by the left shift part of a move. We specify a place by its (position in row,row) coordinates, so that the topmost blob in the left-hand diagram above is at $(4, 5)$.

Note that since we start with a neat pattern each time, the only SW–NE diagonal which can be out of order after the left shift is the bottom one. Thus without loss of generality a move consists of a left shift followed by a (possibly trivial) cyclic permutation of some diagonals at the bottom to restore neatness.

We can run this process backwards to find the predecessors of a given position. A reverse move consists of cycling some SW–NE diagonals to bring a given diagonal to the bottom (the *preliminary cycle*) and shifting the blobs one place to the right, allowing them to wrap round the diagram if necessary. We call the cycle *useless* if, in each column, all the sites in diagonals affected by the cycle contain the same number of blobs, and *useful* otherwise.

REMARK 1. *In performing a reverse move a SW–NE diagonal cannot be brought to the bottom by the preliminary cycle if this would produce a diagram with a blob at the left-hand end of a horizontal row and a space at the right-hand end of the row below.*

PROOF. Otherwise the right shift would lead to a SW–NE diagonal having a blob to the NE of a space, which would be illegal. □

If there is no diagonal such that it can be brought to the bottom and a right shift applied, the position has no predecessor, and is called a *Garden of Eden* position—$[1^3]$ is the smallest such.

By trying all possible preliminary cycles at each stage, we can start to construct a tree of predecessors of a given position. If the position is part of a loop one branch of the tree will bend back to meet the root, otherwise we will obtain a true tree (and in fact if the graph has just one loop of size one (at the root) we will still call it a tree, for convenience). By picking various starting positions we can construct the diagrams shown in Figure 1. In all cases the rightmost branch above each node is the one obtained by using the trivial cycle before the right shift in the backtracking process, i.e. by just applying a right shift.

LEMMA 2. *Δ_k has precisely one predecessor apart from itself.*

PROOF. Consider which SW–NE diagonal could be brought to the bottom by the preliminary cycle to start the backtracking process:

- The bottom one (using the trivial cycle): the predecessor obtained is Δ_k.
- The next one up: the predecessor obtained is $(2, 3, \ldots, k-2, k-1, k+1)$.
- Any other: the cycle is forbidden as in Remark 1. □

Let $\mathcal{P}_k$ denote this unique predecessor $(2, 3, \ldots, k-2, k-1, k+1)$.

PROPOSITION 3. *For $n = T_k$ there is a position whose life (the number of steps required to hit a loop) is $k(k-1)$.*

PROOF. Consider tracing the predecessors of Δ_k up the rightmost branch of its tree—this process consists of repeatedly applying right shifts to the diagram. The first predecessor is $\mathcal{P}_k$, as found in Lemma 2.

Backtracking up this branch will stop if and only if the space at $(0, k)$ in $\mathcal{P}_k$ reaches $(k-1, k)$ as the blob at $(k, k+1)$ in $\mathcal{P}_k$ reaches $(0, k+1)$, for then the next right shift will be illegal, as in Remark 1. After backtracking m moves from $\mathcal{P}_k$, the space will be at (x, k), and the blob at $(y, k+1)$, where

$$x \equiv m \pmod{k} \qquad \text{and} \qquad y \equiv k + m \pmod{k+1}.$$

Thus we want m such that

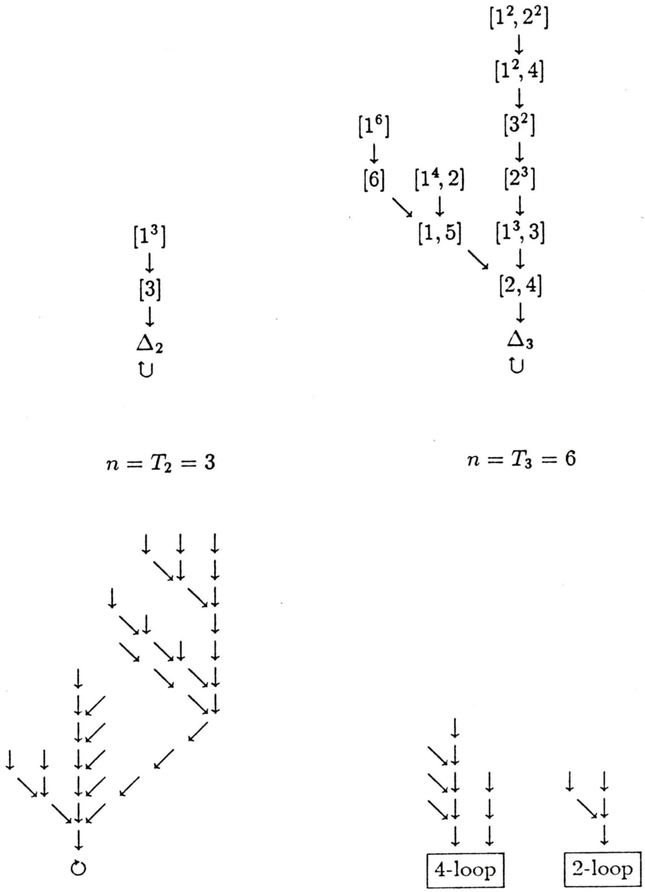
$$m \equiv k - 1 \equiv -1 \pmod{k} \qquad \text{and} \qquad m \equiv -k \equiv 1 \pmod{k+1} \qquad (*)$$

Now

$$k(k-1) - 1 = (k-1)k - 1 \equiv -1 \pmod{k}$$

and

$$k(k-1) - 1 = (k-2)(k+1) + 1 \equiv 1 \pmod{k+1}$$



$n = T_2 = 3$ $\qquad$ $n = T_3 = 6$

$n = T_4 = 10$ $\qquad$ $n = 8$

FIGURE 1. Trees for various values of n.

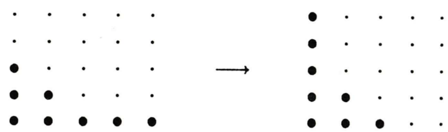
so, since m and $m+1$ are coprime, by the Chinese Remainder Theorem the equations $(*)$ are satisfied if and only if

$$m \equiv k(k-1) - 1 \pmod{k(k+1)}$$

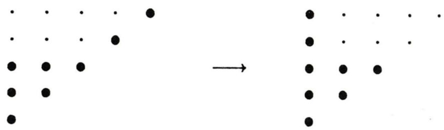
and thus are first satisfied when $m = k(k-1) - 1$.

Thus $\mathcal{P}_k$ has $k(k-1) - 1$ predecessors, so Δ_k has $k(k-1) - 1 + 1 = k(k-1)$ predecessors, the first of which therefore has a life of $k(k-1)$. □

In fact we can say a lot more about this particular chain of predecessors of Δ_k—call it the *main chain*. Given a neat diagram in the unsheared form we can swap its rows and columns (i.e. reflect it in a SW–NE diagonal line through the bottom left blob) to form another neat diagram, which we call the *transpose* of the original diagram. A position $\mathcal{A}$ is the transpose of another, $\mathcal{B}$, if $\mathcal{A}$'s unsheared diagram is the transpose of $\mathcal{B}$'s—we write $\mathcal{A} = \mathcal{B}^T$. Thus $[1^3, 2, 3]$ is the transpose of $[1, 2, 5]$:



Clearly $\mathcal{A} = \mathcal{B}^T \iff \mathcal{B} = \mathcal{A}^T$. In the sheared format, transposition consists of reflecting each horizontal row about its midpoint.



THEOREM 4. *For $n = T_k$ $(k \geq 2)$, and $0 \leq m \leq k(k-1)-1$, the position $\mathcal{A}$ obtained by backtracking m steps up the main chain from $\mathcal{P}_k$ is the transpose of that, $\mathcal{B}$, obtained by backtracking $k(k-1)-1-m$ steps (i.e. $\mathcal{A}$ and $\mathcal{A}^T$ are symmetrically placed in the main chain).*

PROOF. Any (sheared) diagram of a position in the main chain consists of $k-1$ full horizontal rows, a row with exactly one space and a row with exactly one blob. The $k-1$ full rows are unaffected by transposition, and the top two are reflected about their midpoints. Reflecting a row of length r about its midpoint is the same as subtracting the x-coordinate of each of its points from $r-1 \pmod r$.

Now the x-coordinate of row k's space in $\mathcal{B}$ is $k(k-1)-1-m$, hence that of row k's space in $\mathcal{B}^T$ is

$$k-1-(k(k-1)-1-m) \equiv m \pmod k,$$

i.e. the x-coordinate of row k's space in $\mathcal{A}$.

Similarly the x-coordinate of row $(k+1)$'s blob in $\mathcal{B}$ is $k+(k(k-1)-1-m)$, hence that of row $(k+1)$'s blob in $\mathcal{B}^T$ is

$$k-(k+(k(k-1)-1-m)) \equiv k+m \pmod{k+1},$$

i.e. the x-coordinate of row $(k+1)$'s blob in $\mathcal{A}$.

Thus $\mathcal{A} = \mathcal{B}^T$. $\square$

COROLLARY 5. *For $n = T_k$, the position $[1^2, 2, 3, \ldots, k-2, (k-1)^2]$ has life $k(k-1)$.*

PROOF. It is the transpose of $\mathcal{P}_k = [2, 3, \ldots, k-1, k+1]$. $\square$

In the tree for $n = T_4$, above, none of the branches is higher than the main chain—the tree seems to be pruned to this height. This illustrates the conjecture about the maximum life of a position mentioned earlier. An obvious way to prove this conjecture, at least for positions leading to Δ_k, would be to show that, in backtracking from any position, repeated right shifts yield at least as many predecessors as applying a single non-trivial cycle and then applying repeated right shifts (and hence that the main chain is at least as long as any other branch of the tree). Unfortunately this is false: for example for $[1,5]$.

Now for the characterisation of possible loops for general n.

PROPOSITION 6. *Loops consist precisely of those sequences of positions which can be traced by backtracking using right shifts only, without any (useful) preliminary cycles.*

PROOF. Suppose not. Then there is a sequence around which we can backtrack indefinitely, using at least one useful cycle each time we go around the loop.

The bottom SW–NE diagonal of a sheared diagram consists of sites with coordinates of the form $(x-1,x)$. Let a be the smallest integer such that there is a space at $(a-1,a)$ immediately before a right shift (i.e. between any preliminary cycle and a right shift). Note that $a>1$ (for otherwise the diagram can contain no blobs).

Step 1. Consider one of the sites α in row a as it is shifted along the row by the right shifts, and as blobs and spaces are moved vertically in and out of it by the preliminary cycles.

CLAIM: *If α ever contains a space, it will contain a space for ever afterwards.*

PROOF. If a blob replaces a space at α, this must be due to a preliminary cycle. We say that α is *to the left* if it is in one of the first $a-1$ columns $0,1,\ldots,a-2$, and *to the right* if it is in the last column (Figure 2).

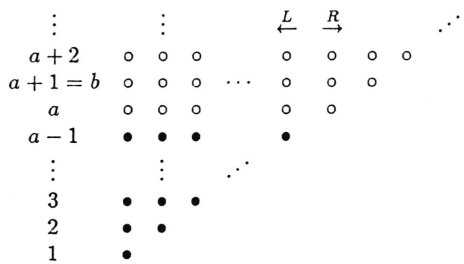


FIGURE 2. o denotes a site which may contain either a blob or a space.

Case 1: α is to the left when the cycle is applied.

A blob is moved into α by the cycle—since the diagram is neat before the cycle is applied, all the sites in the column above α contain spaces and so this blob must come from the site below α. Thus the cycle must bring a space down to a site below α (i.e. below row a), contradicting the minimality of a.

Case 2: α is to the right when the cycle is applied.

By neatness the top site in α's column which is affected by the cycle contains a space. Since α is at the bottom of the column, this space will be brought down to α by the cycle, so the space in α will not be replaced by a blob. □

Step 2: Thus the number of sites in row a containing spaces is monotonically increasing, but must return to the same value each time we backtrack once round the loop, hence must be constant. Hence no blobs can leave row a to be replaced by spaces.

Step 3: Now consider a site β in row $b = a + 1$ as it is shifted right repeatedly.
CLAIM: *If β ever contains a space, it will contain a space for ever afterwards.*

PROOF. If a blob replaces a space at β, this must be due to a preliminary cycle. As before we say that β is *to the left* if it is in one of the first $a - 1$ (sic) columns and *to the right* otherwise.

Case 1: β is to the left when the cycle is applied.

As in Step 1 the cycle must bring a space down to the bottom of β's column, i.e. to a site below row a, contradicting the minimality of a.

Case 2: β is in column $a - 1$ when the cycle is applied.

By neatness the top site in β's column which is affected by the cycle contains a space, which is brought down to row a by the cycle. The cycle must move a blob up from row a to β, so this blob in row a is replaced by a space, contradicting Step 2.

Case 3: β is in column a when the cycle is applied.

By neatness the top site in β's column which is affected by the cycle contains a space. Since β is at the bottom of the column, this space will be brought down to β by the cycle, so the space in β will not be replaced by a blob. □

Step 4: Thus as in Step 2, the number of places in row b containing spaces is monotonically increasing, hence is constant, so no blobs can leave row b to be replaced by spaces.

Step 5: CLAIM. *Row b contains no blobs.*

PROOF. Suppose there is a blob in row b. Since a and b are coprime, in not more than ab steps we will reach a position which has a space at the right-hand end of row a and a blob at the left-hand end of row b (Steps 2 and 4 ensure that the relevant sites will still contain a space and a blob). But then by Remark 1 we cannot backtrack any further. □

Step 6: By hypothesis the backtracking involves at least one useful cycle. By neatness and the minimality of a, in each column which is to the left the part of the column which is affected by this cycle must be full of blobs. Thus the cycle can involve only the bottom two SW–NE diagonals (lest part of row b is forced to contain blobs, contradicting Step 5). Now consider the action of the cycle in column $a - 1$. If it swaps a blob in row a with the space in row b this contradicts Step 2. If it swaps a space in row a with the space in row b then by neatness and legality column $a - 1$ and all the columns to the right of it are entirely empty (contain no blobs) and so the cycle is useless.

Thus the loop can involve no useful cycles. □

Call a horizontal row of a diagram *mixed* if it contains both blobs and spaces.

THEOREM 7. *A diagram is part of a loop if and only if it contains at most one mixed row.*

PROOF. By Proposition 6 backtracking round any loop involves only right shifts. Suppose a diagram has two mixed rows c and d, with $2 \leq c < d$. By neatness and legality row $c+1$ must be mixed. As in Step 5 of the proof of Proposition 6, since c and $c+1$ are coprime backtracking must fail in at most $c(c+1)$ steps.

Conversely if a diagram has fewer than two mixed rows, we can backtrack indefinitely using only right shifts, since the situation warned of in Remark 1 can never arise, and thus the diagram is part of a loop. □

COROLLARY 8. *For $n = T_k$, the only loop is the one-step loop consisting of Δ_k.*

PROOF. No neat diagram with T_k blobs can have precisely one mixed row (if it does, suppose it has r full rows, then by neatness and legality, row $r+1$ must be the mixed row, but if s is the number of blobs on this row, $T_r + s = T_k$ has no solutions for r and s with s between 0 and $r+1$ exclusive). Thus by Theorem 7 the only diagrams which can be part of a loop are those with no mixed rows, and the only one of this form is Δ_k. □

How many loops of each size are there, for general n? If k is such that $T_k \leq n < T_{k+1}$ we form a loop diagram by filling the first k horizontal rows and placing the remaining $n - T_k$ blobs on row $k+1$.

Let $u = k+1$, $v = n - T_k$, and let $\lambda(u, v, s)$ denote the number of distinct loops of size u/s obtained by placing the v blobs in the u sites of row u (so that for a position in one of these loops s is the order of symmetry of row u with respect to horizontal shifts).

LEMMA 9.

$$\lambda(u, v, s) = \begin{cases} \lambda(u/s, v/s, 1) & \textit{if } s \mid (u, v), \\ 0 & \textit{otherwise,} \end{cases}$$

where (u, v) denotes the greatest common divisor of u and v.

PROOF. If $s \mid (u, v)$, let $u' = u/s$, $v' = v/s$. There is a one-to-one correspondence between ways of placing v' blobs in u' sites so that the resulting pattern has symmetry of order 1, and ways of placing v blobs in u sites so that the resulting pattern has symmetry of order s (since in the second case the placement of v' blobs in the leftmost u' sites completely determines the placement of the remaining blobs).

If $s \nmid (u, v)$ it is clear that no placement of v blobs in u sites can have symmetry of order s. □

Given this result, it is enough to be able to calculate $\lambda(u, v, 1)$ for arbitrary u and v. Note that unless $v = 0$ there is always at least one loop of size u—place the v blobs as far to the left of row u as possible.

PROPOSITION 10.

$$\lambda(u, v, 1) = \frac{1}{u}\left(\binom{u}{v} - \sum_{1 \neq s \mid (u,v)} \left(\frac{u}{s}\right) \lambda(u/s, v/s, 1)\right)$$

PROOF. Given u and v, each of the $\lambda(u, v, s)$ loops of patterns with symmetry of order s contains (u/s) of the $\binom{u}{v}$ possible placings of the blobs, so

$$\begin{aligned}\binom{u}{v} &= \sum_{s|(u,v)} \left(\frac{u}{s}\right) \lambda(u,v,s) \\ &= \sum_{s|(u,v)} \left(\frac{u}{s}\right) \lambda(u/s, v/s, 1) \\ &= u\lambda(u,v,1) + \sum_{1\neq s|(u,v)} \left(\frac{u}{s}\right) \lambda(u/s, v/s, 1)\end{aligned}$$

and the result follows. □

Although this may seem a poor result, being only a recurrence relation, in practice the fact that $u \sim \sqrt{2n}$ and v is even smaller means that (u,v) is small and so there are not many terms on the right-hand side.

For example, the game played with a standard deck of 52 cards has 12 loops, each of length 10, since in this case $u = 10$ and $v = 7$ are coprime. $n = 1990$ is equally dull, since again $u = 63$ and $v = 37$ are coprime and so the game has 5,669,444,052,290,067 loops, all of length 63. Adding a joker to the standard deck is more interesting, yielding a game with $(u,v) = (10,8) = 2$, and hence 4 loops of length 10 and 1 of length 5, while taking $n = 288$, say, has $(u,v) = (24,12) = 12$ and hence yields a game with a 2-loop, a 4-loop, three 6-loops, eight 8-loops, seventy-five 12-loops and 112,632 24-loops.

REFERENCE

[1] "Mathematical Games" in *Scientific American*, August 1983.

The Creative Archimedean At Work

An old tradition of the Archimedeans is the annual Puzzle Hunt, where candidates are asked to run around Cambridge solving puzzles. A more recent tradition decrees that one question require an answer in the form of an essay. Candidates must write about "Life in Cambridge" in no fewer than 150 words, or 60 words if they write in verse. This year's overall winner, David Moore, took advantage of the latter clause to gain time by writing what is undoubtedly the worst effort in the poetical line to have come my way for as long as I remember. He has paid me large sums of money not to reprint this. A prize is also awarded for the best essay, and this went to Michael Aird for his charmingly vomitory prose offering:

> Every morning I wake up to the sound of my alarm clock ... I throw up in the sink as a result of the coughing fit I have to endure. I go to breakfast, throw up on my food, and run back to my room, throwing up in Trinity Street.
>
> I go to lectures next. I throw up because they are so complicated that throwing up gives me something to do that I can understand.
>
> Lunch is good. I throw up. ...

Perhaps I should add by way of reassurance that, in everyday life, Michael is a less emetical character than his prose might lead one to imagine.

Call My Bluff

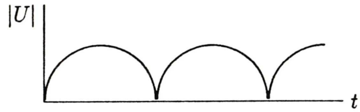
Compiled by Stephen Turner

The following are all based on definitions given at the Inter-University Call My Bluff competition held in the OCR, Trinity on 3rd December 1989. Questions 1–7 were set by the Cambridge team, and question 8 by the chair and assistants. The answers, and a short report of the competition, may be found on page 90.

For each of the following words, which of the three given definitions is correct?

1. Bang-Bang type

(a) A class of theories, suggesting that the evolution of the universe has a cyclic nature, consisting of alternate "big bangs" and "big crunches".



Any such theory is said to be of the 'Bang-Bang type'.

(b) A type of control function u such that each component u_j of u takes only the values $+1$ and -1.

(c) A group is of finite (respectively countable) bang-bang type if it can be expressed in generator relation form, with finitely (respectively countably) many generators (a_i), $i = 1, 2, 3, \ldots$, and relations of the form $a_i^{i!} = a_j^{j!}\ \forall i, j$.
The name "bang-bang" refers, of course, to the factorials in the exponents.

2. Quadratix

(a) The curve $y = x\cot(\frac{\pi x}{2})$

(b) A character from the book *Asterix in Greece.*

(c) Any pair of simulataneous equations of the form

$$ax + by = c$$

$$cx + dy = a$$

with $(ad - bc) = 0$ (a quadratix has a solution if and only if $a = \pm c$)

3. Contorted Fractions

(a) A two-player game invented by John Conway, in which the legal moves for each player involve changing one of a set of numbers so that its denominator is reduced.

(b) A generalization of the notion of a continued fraction, conceived and named by John Conway.

(c) In his book *On Numbers and Games*, John Conway considers the surreal numbers. The contorted fractions are the surreal counterpart to the rationals in $\mathbb{R}$.

4. Syrtis

(a) Syrtis is pronounced "Seer-tis". A syrtis is any curve of the family

$$(x^2 + y^2)^2 - 2a^2(x^2 - y^2) = 0$$

which may be more familiar in polar form

$$r^4 - 2a^2r^2 \cos 2\theta = 0$$

(b) Syrtis, pronounced "Sire-tis", is the character representing the empty set, $\varnothing$, commonly mistaken for phi, ϕ.
(c) Syrtis is in fact pronounced "sur-tis" and has nothing to do with maths.—it is a word Milton used meaning quicksand.

5. Suanching-Shihshu

(a) An early Chinese mechanical computing machine.
(b) The Suanching-Shihshu method is a way of renormalising in superstring theory, to remove the divergent integrals that would otherwise arise.
(c) Suanching-Shihshu (plural) are ten books on arithmetic, written in China between the second century BC and the sixth century AD.

6. Bornological

(a) In topology, a bornological space is a locally convex space, such that every absolutely convex set that absorbs every bounded set is a neighbourhood of 0.
(b) A map from $\mathbf{C}^n$ to $\mathbf{C}^n$ is said to be bornological if and only if it has exactly one fixed point.
(c) A QM scattering is Bornological if the Born-Oppenheim approximation yields an answer for the energy which is correct to $O(\frac{\mu^2}{M^2})$ where μ is the reduced mass of the system and M is the total mass of the system.

7. Doetsch's Three-Line Theorem

(a) In 1986, Professor J. G. Thompson completed the classification of finite simple groups with a theorem whose proof was so complex that it covered several densely printed pages of the *Proceedings of the American Mathematical Society*. However, in 1988, Doetsch realised that the necessary result followed as a (relatively!) simple generalisation of a theorem due to Feit, and so the proof can now literally written on only three lines. For this reason, the theorem is now known, somewhat misleadingly, as Doetsch's Three-line Theorem.
(b) Let $L(\sigma) \equiv \sup_{b\in\mathbf{R}} |f(\sigma + ib)|$, $(\sigma \in \mathbf{R})$ for a function f that is bounded and regular in a vertical strip $\alpha < \mathrm{Re}(z) < \beta$ in $\mathbf{C}$. Then $\log L(\sigma)$ is a convex function of σ in $\alpha < \sigma < \beta$.
(c) Doetsch's three-line theorem is a theorem in spherical geometry. It states that given a sphere radius r, and a triangle on its surface whose angles are of magnitude a, b, c, the area of the triangle is $r^2(a + b + c - \pi)$.

8. Eikonal

(a) The eikonal equation is

$$\frac{\partial\psi}{\partial x_i}\cdot\frac{\partial\psi}{\partial x^i}=0,$$

where ψ is the wave phase. It is the fundamental equation of geometrical optics.

(b) In QM, an eikonal is an expression of the form

$$\langle L_1 M_1 | J_i J_j | L_2 M_2 \rangle$$

where **J** is the total angular momentum operator, and L_i, M_i are the eigenvalues of the individual angular momentum acting on the i^{th} particle.

9. Who said ...

"Mathematics ... possesses not only truth, but supreme beauty"?

(a) Isaac Newton.
(b) Bertrand Russell.
(c) G. H. Hardy.

An Important New Result

It is possible that some readers of *Eureka* are acquainted with the field of *Voidology Theory*, a branch of mathematics whose full potential is still only beginning to emerge. The seminal paper is of course [1]. Many people have stated that voidology has no application to any other branch of mathematics, but they will now have to reconsider for voidologists recently came up with an entirely unexpected proof of the well-known *Null Hypothesis*. This is now elevated to the status of a theorem, and a sketch of the proof is appended; readers acquainted with the basic concepts of voidology shoud have no difficulty in following it:

THEOREM (*The Null Theorem*).

PROOF. □

Some voidologists are now said to be working on a proof of the Generalised Null Hypothesis, generally considered to be an altogether more difficult nut to crack.

REFERENCE

[1] Mitchell, W. J. R., "Voidology Theory", *Eureka* **44** (1984) pp. 25–29.

Perestroika in the Maths Faculty

Heather Mendick

Every thirty years or so outside influence and educational considerations combine, resulting in the decision to restructure the Cambridge Mathematics Tripos. The last overhaul was in the early 1960s when more modern material was introduced into the first year of the course, making the standard route to Part II at the end of three years, rather than two as previously.

There are two main problems that have provoked the present changes. The first, which is external, is the movement away from specialisation at sixth form. The complete rejection of the Higginson report by the government in 1988 has in many ways made matters more difficult, since it leaves the situation ambiguous. It is expected that a large proportion of schools (especially those in the State sector) will adopt the spirit of the proposals, consistent with current educational philosophy, that sixth form studies should offer students greater breadth than (for example) the traditional double maths. and physics course encouraged by colleges for entry to Cambridge. Moreover, even if schools are willing to offer Further Maths., they are often unable to do so, because of the now desperate shortage of qualified maths. teachers. Thus if Cambridge continues to insist on such exacting qualifications it will be drawing on a declining pool of mathematicians and will lose a lot of talented but less well-prepared people to other universities.

The second problem is internal and represents the feeling of many department members that we are failing a large minority of very able students who do not succeed within the Tripos. Many of the "tail of poor performers" do not finish the course (maths. has a 20 per cent dropout rate), and of those who do, many fall into the third class. In recent years external examiners have raised doubts as to whether, on the evidence of their Part II examinations, these people would have been passed at other British universities.

With these objectives in mind a Committee on Maths. in Schools (the Crighton Committee) was created by the faculty board to develop proposals for consideration by the departments and the faculty. This job took about six months and the final report, which was circulated over the summer, contained three schemes (for pseudo-psychological reasons—apparently three is a good number to have); however, it was essentially a majority report accompanied by two minority reports which were eventually and predictably rejected.

Although two out of three of the schemes listed titles of courses, this was simply to help people to visualise how they would work. It must be stressed that what was being put forward in each case (and what I believe that the final decision was based upon) was a philosophy and a structure rather than details of particular courses. We all recognised:

> that all courses will have to be completely re-thought in content, and that significant departures from current practice will probably have to be made in teaching style. In some cases it will be necessary to plan courses for two years as a single unit (and possibly related, further, to Part III). It may be

> necessary and desirable to prepare comprehensive notes, examples, and other material with far more cohesion, thoroughness and detail than previously.
>
> It may also be necessary, and desirable, to supplement some university and college teaching with examples classes.
>
> *Crighton Report*, page 2.

The least popular proposal, possibly because of its similarity to the eventual winner, was the "Split Parts IB and II Tripos". It consisted of a common first year accessible to single "A"-level students, culminating in a classified Part I examination.

> The ... course would be similar in scope to the present Part IA but slipped in time by perhaps half a term and reduced in content. A completely new course would have to be written, and much teaching throughout the year would have to be of a less formal character than now (closer to school patterns, and involving use of examples classes).
>
> *Crighton Report*, page 9.

The year would then split into parallel "General" and "Special" streams, the distinction being based on students' interests rather than abilities, with Part II General being aimed mainly at those not intending to continue with maths. after Part II. It was rejected because the the common first year took no account of the very different levels of familiarity with maths. students would have on starting the course, and also because the split at the end of the first year, although not fatal in that good Part II General students could proceed to Part III, was felt to be much too early.

The other scheme to be rejected was split entry, based on the "underlying conviction" that:

> ... most students with one "A"-level in mathematics cannot initially be taught with most of those with two "A"-levels, nor can they be examined together in the first year without the expertise of the latter crushing the spirit of the former. This means there must be no Tripos exam., no public classing, in the first year. By the end of the second year the clever but ill-prepared must have caught up with the well-schooled but not always capable students, and all can know they are doing *the* Mathematical Tripos.

Although some people wrote in support of this scheme, many strongly opposed the idea of separating people on entry into an A and an AA stream. It was also felt that the "tail" would not be eliminated simply by tinkering with the present Parts IB and II by removing a couple of optional courses from the former and adding three or four more user-friendly courses to the latter. It had the added disadvantages of having no classed Part IA exam. at the end of the first year, but instead a second-year exam. on two years' work; and of a summer vacation course for the A students with financial and residence implications.

The scheme that was adopted at the maths. faculty AGM in November 1989, despite some protests from members of "a certain riverside college", differed fundamentally from the second idea outlined above in that, while it was designed with the well-prepared in mind, it was aimed at the lowest third of students. This was a positive decision which should send the right message to schools and indicate how genuine is Cambridge's decision to accomodate single "A"-level mathematicians within the Tripos. The structure shown in Table 1 differs in minor details from that given in the Crighton document since in the process of implementation it became necessary to modify it. The course is being slowed down with more introductory material (for example on methods of proof) being introduced, and with a combined pure and applied approach to the first

TABLE 1. THE SPLIT PART II TRIPOS.

	Term	Core Courses	Pull Forward	Options
First year	1	C_1, C_2, C_3, C_4	—	—
	2	C_5, C_6, C_7, C_8	P_1	—
	3	D_1, D_2, D_3, D_4	P_2, P_3	—
	Part IA Exam on C_1–C_8			
Second year	4	C_9, C_{10}, P_1, P_2	P_4, P_5	O_1, O_2
	5	$C_{11}, C_{12}, P_3, P_4, P_5$	—	O_3, O_4, O_5, O_6
	6	D_1, D_2, D_3	—	—
	Part IB exam. on C_9–C_{12}, P_1–P_5, D_1–D_4			
	"O" Paper on two out of O_1–O_6 (non-Tripos and optional)			

Part II Option "A"

Anticipated for the 60–70% of students whom Tom Kõrner has dubbed "the student in the street". It will consist of courses designed for those for whom Part II is their last experience of University maths. and will hopefully leave them "feeling warm all over".

For practical reasons it is necessary that O_1–O_6 should be available in this option.

Part II Option "B"

For the remaining 30–40% of students, dubbed "high fliers" by Tom Kõrner. Its courses will be of a more technical nature for those intending to do research—a sort of Part IIIA. Entry will be controlled by Directors of Studies but will normally be restricted to those obtaining a first in the IB exams or a good second in Part IB and a pass in the optional paper.

There will be no mixing of courses between the two Part IIs and firsts will be awarded in both. It will be possible to proceed from option "A" to Part III, which remains unchanged by the proposals, just as Part III is now accessible to those who have graduated from other universities.

NOTES:

$C_n = n^{\text{th}}$ core course, lectured once (unless double lecturing forced by lack of space)

$P_n = n^{\text{th}}$ core course given also in pulled forward version

$D_n = n^{\text{th}}$ 12-lecture core course taken in third *or* sixth term (except D_4, a computing course to familiarise people with PWF, Pascal and numerical techniques which remains a *"D"* course by default)

$O_n = n^{\text{th}}$ optional course.

two terms, which should put an end to idiosyncrasies of the present IA course such as three different first-term courses teaching vectors in three different ways. A list of the courses is given in Table 2. There will be a classified IA exam. at the end of the first year on the material covered in these two terms, to enable people to change subject easily. However, it will be fairly straighforward and so will enable the third term to be used for teaching to a greater extent than at present. By the end of two years, everyone will have covered a compulsory block of material, which in Analysis goes up to Complex Variable and in Methods covers Vector Calculus, solution of differential equations, Sturm-Liouville, the transform calculus, the residue theorem, calculus of variations and tensors. Students can progress through the Part I course at a faster rate by "pulling forward" courses from later terms. Thus those who find the IA material elementary can supplement it in the second term with Linear maths. (P_1), which is

TABLE 2. LIST OF COURSES.

Course(s)	Lectures Per course	Subjects
C_1/C_2	24	Algebra and Geometry in $\mathbf{R}^2$ and $\mathbf{R}^3$
C_3	24	Discrete mathematics
C_4	24	Differential equations
C_5/C_6	24	Div, grad and curl; Groups, transform calculus
C_7	24	Probability
C_8	24	Dynamics
D_1–D_3	12	Numerical analysis, geometry, optimisation
C_9	24	Analysis
C_{10}	24	Mathematical methods
C_{11}	16	Statistics
C_{12}	16	Further Analysis
P_1	24	Linear Mathematics
P_2	16	Electromagnetism
P_3	16	Complex methods
P_4	16	Further Algebra
P_5	8+16	Special Relativity and Quantum Mechanics
O_1/O_2	24	Classical mechanics, Markov Chains
O_3–O_6	24	Analysis, Algebra, Fluids, Dynamical Systems

taken in its standard position in term 4. P_1 can then be replaced by pulling forwrd Further Algebra (P_4) from term 5 into term 4. Allocating P_2 and P_5 to Theoretical Physics creates a similarly consistent "pull forward" rate in applied maths. In order to avoid unwise premature decisions, no pull forward courses have been scheduled for the first term. The slots freed in some students' second year curriculum (by pulling forward courses) are filled by optional courses, which are examined with the IB exam in a separate, non-Tripos paper. These optional courses are important because of the split structure of Part II (see table): they fulfil the dual function of courses in option "A" of Part II, and prerequisites for courses in option "B". This creates a conflict between the need for them to be suitable for students concluding their academic mathematical studies, and to include results necessary for technical Part II courses. While the first objective is the primary one, it is hoped that their two rôles can be reconciled to create interesting, relevant courses.

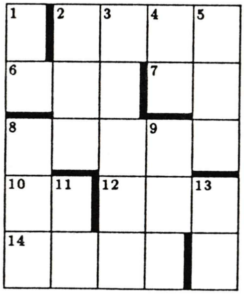
The Crighton Committee has long since been disbanded and replaced by the Tripos Committee, in charge of implementation, and a glut of scheduling committees, the most important of which, the First Year Core Course Committee, has nearly completed its work on the content of C_1–C_8. The work of the others is about to begin and a document defending the allocation for subjects to C, P, D and O slots should by now have been released. If all goes according to schedule the restructured Tripos should be in place in most of (if not all) its glory for the 1991 intake, and we hope to have the new exam. structure passed by the general board next term as well as having Part I schedules in a provisionally complete state in time for a special joint meeting of the departments in June. After the disaster of the flexi-speed scheme, everyone realises how important it is that we get it right this time.

Problems Drive 1990

A. Frazer Jarvis and The Graham Nelson

1. The answers to the following crossnumber satisfy:
(a) No answer begins with a zero.
(b) Within each answer, no digit appears more than once.
(c) The sum of the digits in each answer is given as the clue.

Complete the grid:



Across

2.	15
6.	15
7.	14
8.	28
10.	15
12.	12
14.	7

Down

1.	11
2.	12
3.	24
4.	8
5.	9
8.	15
9.	16
11.	12
13.	14

2. The mad scientist, Professor Owen St John, a very aged and hirsute person, has constructed an infernal machine, which he calls the $Z0$. The $Z0$ holds one number stored in it at all times, which we shall call A, and which has some value from 0 to 15. The $Z0$ is started by pressing a button, which sets A to 0 and starts it working. When it is working, it performs one task at a time, according to some arcane internal program. Its possible tasks are:

DAA	Double A (modulo 16)
SUFA	Subtract 4 from A (modulo 16)
GEYO	Display a green light if A is even, or a yellow light if A is odd.
INP	Ask the operator to enter a number (from 0 to 15) which is then added to A (modulo 16)
END	Immediately stop

At the end of performing DAA, SUFA or INP, 1 is added to A (modulo 16) automatically. DAA, SUFA and GEYO take 1 minute each, while INP takes as long as you take to enter a number.

Unfortunately, there is a bug in the $Z0$, and if A ever has either the value 8 or 11, then the $Z0$ explodes and kills the operator (at which point the Professor will say *"Le compte est bon"*).

The Professor munches a peppermint and, brandishing a deadly ray gun, forces you to press the button. Two tense minutes pass, then a green light shows for a minute and then two more minutes pass. The $Z0$ then asks you to enter a number. The Professor tells you that the next instruction will be END.

What number do you enter?

3. The tables of Tripos performance are, of course, published every year. This year's list of the top ten colleges compares with last year's in the following ways:

(a) Seven colleges are in both lists.

(b) The college which finished fifth this year was also fifth last year.

(c) Each of the other colleges which appear in both lists have improved their position, and all by a different number of places; the largest gain was achieved by Selwyn College, which came first this year.

(d) There are no ties.

In how many ways can this happen?

(HINT: $1+2+3+4+5+6 = 3 \times 7$.)

4. Every year, the Cantabrigian Conservatory of Modern Art takes on seven or so young painters out of about seventy hopeful artists. By the charming old customs of the Conservatory, the lucky seven are chosen by being examined on 31st May on their ability to recite randomly chosen pages from the (10000 page) Cantabrigian District Telephone Directory (1957 edition). Candidates may begin work when the year starts on 4th October, but will not be able to do any work on the actual day of the exam. Each day spent working allows the memorisation of $23 - 3n$ pages, where it is the n^{th} day in a row spent working (thus on the initial day, 20 pages may be learnt, on the next 17, and so on). So it is obviously advisable to take some days off work to "recover". What is the largest number of pages it is possible to commit to memory before the exam?

(NOTE: the exam. year is a leap year).

5. Within a regular tetrahedron, there lies a sphere. A smaller regular tetrahedron lies inside the sphere. If the large tetrahedron has volume of one cubic unit, how large may the small tetrahedron be? If it is a cube which lies inside the sphere (instead of the small tetrahedron), what is the largest volume it may have?

(HINT: for the second part, use the first part, and that the volume of a tetrahedron of side x is

$$x^3 . \frac{\sqrt{2}}{12} \qquad)$$

6. You are visiting the World Psychiatrists' Congress, where you are to give a paper on "Mental Diseases amongst Perfect Logicians". To demonstrate, you are going to exhibit five patients, A, B, C, D and E, sent to you by local asyla. Tragically, these perfect logicians are crippled by the following disorders:

A: can only lie, but is able to choose whether to speak or not, and will always try to be as helpful as possible to people trying to discover his identity;

B: can only tell the truth (he is in an asylum since, mysteriously, he tends to say "Swooshh!" when he walks past things);

C: is entirely mad and so no inferences may be drawn from his conversation at all;

D: is a manic depressive, and will either respond to questions in a positive (not necessarily truthful) way to ingratiate himself with the questioner, or falls into a moody silence;

E: is convinced he is a duck, and says "Quack!" occasionally.

Unfortunately, you are unaware of which one is which, and although all five of your patients have read the abstract of your paper and know all about their respective disorders, they are also unaware of one another's identities. We shall call your patients V, W, X, Y and Z.

You ask them, "Which one of you is A?"

Three of them speak at once; X says "Me!", Y says "You are!" (pointing at you) and Z says "Quack!". V and W are silent. Deciding to diagnose them, you ask "What is the first word that comes into your head?"

There is a short pause, then X and V say "Quack!", while W says "What!" Z says "Jellyfish!" and Y is silent.

Which is which?

7. The game "Eureka-Qarch" is played thus:

The players count $(0, 1, 2, \ldots)$ in turn. Thus if the players are A, W, C then A begins with 0, W follows with 1, then C with 2, A continues with 3 and so on. However:
If a player is to say a number divisible by 3, he or she must say EUREKA (E);
If a player is to say a number divisible by 7, he or she must say QARCH (Q);
If a player is to say a number divisible by both, he or she must say EUREKA-QARCH (E-Q);
If a player is to say a number divisible by neither, he or she must say NEITHER (N).

So the sequence would normally begin:

0	1	2	3	4	5	6	7	8	...
E-Q	N	N	E	N	N	E	Q	N	...

If a player makes an error, he or she retires from the game; the next player continues with the next number. The game ends when one player remains.

Under the Charter of Charemaidens' College, the Prime Minister of the day normally chooses the next Master. The Fellows, who want to elect a mathematician, have found an ancient Statute allowing them to use this game instead to select the Master.

I arrived late to the election; four candidates were standing and all were still playing. I heard:

N E N Q N N N E Q N E

at which point one candidate won.

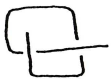
If the winner was the starting player (with 0 E-Q), and the game ended at the first possible opportunity given the above information, what was the number corresponding to the last turn?

8. The Not Knot game is played by two players, Under and Over, who alternately draw on a piece of paper. The initial move is by Over, who just draws a straight line:

Then Under must continue one end and pass it under exactly one drawn line. Now consider what you would get if you joined the two ends together by a straight line going over everything in the drawing. If this is knotted (that is, not just a loop, but a loop with a knot tied in it) then Under loses. Otherwise, Over then continues an end and passes it over one already drawn line. Over loses if the loop you'd get by joining the ends with a line under everything has a knot in it. So for example, Over does the line

Under loops it

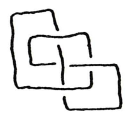
Over does this



and loses, since if you join the ends you get

which is knotted.

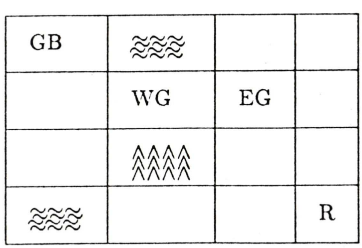
The question is: give a legal game which ends with one player losing with the figure-8 knot, which looks like this:



9. In the following entirely imaginary continent, the regions on the map denote countries which, on September 1st, all have one-party state dictatorships. During the course of September, Ruritania (R) has a revolution and becomes free and democratic.

Due to Politburo and Cabinet meetings being held on the 1st of each month in each state, every neighbouring state (where diagonal squares are considered to be adjacent) becoming democratic during the previous month causes the state in question to become democratic with probability one-half, as from the 2nd of that month. However, East and West Gemland (EG, WG) will, as soon as *either* one becomes free, immediately reunify into one large state, which will be democratic.

What is the probability that Grand Burgundy (GB) will be democratic before the year is out?



Mountains Sea

There are no countries in squares occupied by mountains or sea.

10. Match the following "famous" mathematical constants to the approximate numerical values given below:

A. Brun's constant, the sum over all prime pairs $(p, p+2)$ of $\dfrac{1}{p} + \dfrac{1}{p+2}$

B. Euler's constant, the limit as $n \to \infty$ of

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n$$

C. Skewes' number, an upper bound on the x for which $\mathrm{Li}(x) - \pi(x)$ first changes sign, where

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}, \qquad \pi(x) = \text{no. of primes } \le x.$$

D. Feigenbaum's number, the limit of the speed at which period doubling occurs in many dynamical systems.

E. Artin's constant $= \displaystyle\prod_{p \text{ prime}} \left(1 - \frac{1}{p(p-1)}\right)$.

F. 23/71.

G. π^e.

H. e^π.

I. $\pi/\sqrt{18}$, conjectured to be the closest possible packing of identical spheres.

J. $1/\pi$

K. Linnik's constant, defined as the smallest L such that for every sufficiently large d, $p(d) < d^L$ where

$$p(d) = \max \left\{ \begin{matrix} \text{smallest prime in the AP } a, a+d, a+2d, \ldots \\ \text{s.t. } 0 < a < d \text{ and } \mathrm{HCF}(a,d) = 1\} \end{matrix} \right\}$$

a 0.74048049
b 4.669201660910
c 0.31830988618379067153776752674502872406891929148O
d 23.140692632779269005729086
e 0.373955813619202288054728054346416415111629249
f Between 1 and 13.5 but conjectured to be 2
g 22.459157718361045473427152
h 0.577215664901532860606512090082402431
i 0.32394366197183
j About $10^{10^{34}}$
k 1.90216054

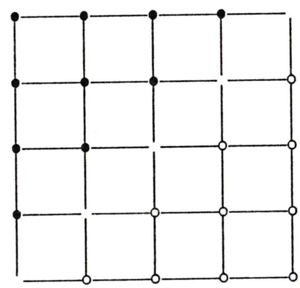
For a small bonus, which one of these is also known as Mascheroni's constant?

11. What is the next number in the following series?

(a) $3, 7, 10, 17, 27, 44, \ldots$
(b) $3, 7, 12, 18, 26, 35, \ldots$
(c) $3, 7, 13, 19, 29, 37, \ldots$
(d) $3, 7, 11, 20, 28, 40, \ldots$

12. The game of "Go" is played on the *intersections* of an 18×18 grid (including those at the edges of the grid). Coincidentally, so is this problem. The board is almost full with only the diagonal running from bottom left to top right being free of "stones", and black stones on all the intersections above and to the left of this diagonal, but white stones on all those below and to the right. If one "move" consists of moving a stone by one unit horizontally or vertically to an unoccupied square, what is the least number of moves needed to fill the upper portion with white stones and the lower with black?

I.e., at the start we have an 18×18 version of this 4×4 grid:



and we have to make the upper triangle full of white pieces and the lower one full of black pieces.

THE RULES

Pairs played as teams and had to answer twelve questions in one and a quarter hours. Every five minutes they were given a new question, and the questions were removed from them after 10 minutes; at the end teams were given 10 minutes to invent any more plausible answers: guesswork is usually the decisive factor. Each question carries an equal weight of marks, making the scoring system complex and incomprehensible (this year's denominator was 360). The winners receive a bottle of port and are expected to set the questions the following year.

Beyond the Vanishing Point

Ian Stewart

What is the area of a circle? Why?

Countless generations of schoolchildren have had the answer to the first question drummed unto their heads until they can recite it in their sleep:

$$A = \pi r^2.$$

The other formula involving π which every schoolchild knows is $2\pi r$, which gives the length of the circle's circumference.

Because it is transcendental, π is a little tricky to handle. In particular, we may ask: how do we now that π is the *same* number in both cases? Experimental measurements on the area and circumference of circles might perhaps show that the two values agreed to, say, nine decimal places. That's good enough in practice; but it won't help with the theoretical problem "Is the agreement *exact*?" There are plenty of coincidences in mathematics. In particular, π and 355/113 agree to nine decimal places, but that doesn't prove *them* equal.

Either formula can be taken as a definition of the number π. But once we've chosen which, it is necessary to prove that the other is consistent. Specifically, we can write the area πr^2 as $(\pi r)(r)$, and derive a result that does not mention π at all: *the area of a circle is the same as the area of a rectangle whose base is half the circumference and whose height is the radius.* That may seem a bit of a mouthful, but no worse than many a statement found in a geometry book; and it is just what the doctor ordered, because it will prove in all logical rigour that there is only one π, not two. And in fact, it looks as if it ought to be fairly easy.

It isn't, of course: in mathematics it's always the things that look easy that turn out to be downright impossible. But we can get very close ...

The intended "proof" is another old friend. Slice up the circle radially, like cutting up a pie (no pun intended), into lots of thin pieces. Arrange the pieces next to each other in a row, pointing alternately North and South, edge to edge. The result will be a slightly wobbly rectangle. Its height is just the radius r. The original circumference of the circle has been distributed along the top and bottom edge; so each has length half the circumference. So the rectangle's sides are approximately equal to:

(1) the radius,

(2) half the circumference.

Because chopping the curve up doesn't change its area, this nearly does the job. The only snag is that word *approximately.*

If you slice the pie ever more thinly, the error becomes smaller and smaller. But for any finite number of slices, error there is: it never disappears altogether.

Suppose, however, that it were possible to slice the pie into infinitely many infinitesimally thick pieces. Then the error would vanish, and the wobbly rectangle would be wobbly no more. That would solve the whole problem. But is it a legitimate argument?

INFINITESIMALS versus EPSILONTICS

As far as we know, the first people to ask questions about the proper use of logic were the ancient Greeks, although their work is flawed by modern standards. And in about 500BC the philosopher Zeno of Elea invented four famous paradoxes to show that infinity was a dangerous weapon, liable to blow up in its user's hands. Even so, the use of "infinitesimal" arguments was widespread in the sixteenth and seventeenth centuries, and formed the basis of many presentations of (for example) the calculus. Indeed it was often *called* "Infinitesimal Calculus". The logical inconsistencies involved were pointed out forcibly by Bishop Berkeley in a 104-page pamphlet of 1734 called *The Analyst: A Discourse Addressed to an Infidel Mathematician.* The trouble was, calculus was so useful that nobody took much notice. But, as the eighteenth century wore on, it became increasingly difficult to paper over the logical cracks. By the middle of the century, a number of mathematicians including Augustin-Louis Cauchy, Bernard Bolzano and Karl Weierstraß, had found ways to eliminate the use of infinities and infinitesimals from the calculus.

The use of infinitesimals by mathematicians rapidly became "bad form", and university students were taught rigourous analysis, involving virtuoso manipulations of complicated expressions in the Greek letters ε (epsilon) and δ (delta) imposed by the traditional definitions. There is even a colloquial term for the process: *epsilontics.* Despite this, generations of students in Engineering departments cheerfully used the outdated infinitesimals; and while the occasional bridge has been known to fall down, nobody to my knowledge has ever traced such a disaster to illogical use of infinitesimals.

In other words, infinitesimals may be wrong—but they work. Indeed, in the hands of an experienced practitioner, who can skate carefully round the thin ice, they work very well indeed. Although the lessons of this circumstance have been learned repeatedly in the history of science, it took mathematicians a remarkably long time to see the obvious: that there must be a reason why they work; and if that reason can be found, and formulated in impeccable logic, then the mathematicians could use the "easy" infinitesimal arguments too!

It took them a long time because it's very hard to get right. It relies on some deep ideas from mathematical logic that derive from work in the 1930s. The resulting theory is called *Nonstandard Analysis*, and is the creation of Abraham Robinson. It allows the user to throw real infinities and infinitesimals around with gay abandon. Despite these advantages, it has yet to displace orthodox epsilontics, for two main reasons:

- The necessary background in mathematical logic is difficult and, except for this one application, relatively remote from the mathematical mainstream.
- By its very nature, any result that can be proved by nonstandard analysis can also be proved by epsilontics: it's just that the nonstandard proof is usually simpler.

Isaac Newton hit the same problem when he first introduced calculus: he actually presented the arguments in a "disguise" so that they looked like the usual geometrical proofs. In the end the calculus won, because it was a simpler and more powerful technique: on really complicated problems you *could* cobble up a geometric proof, but it would be too messy for anyone to understand what it was about. Furthermore, using just geometry, nobody would ever have thought of the proof in the first place.

Despite a certain degree of resistance from the more conservatively minded mathematicians, Nonstandard Analysis is beginning to gain popularity. This process, begun on largely aesthetic grounds, may soon speed up considerably. Because there is growing evidence that Nonstandard Analysis may prove an important tool not just in pure mathematics, but in applications. Applied mathematicians are forever making use of the fact that certain numbers in their equations are "small". The new idea is to go the whole hog, and make those quantities become true infinitesimals. The result is a new tool for both qualitative and quantitative analysis.

EXPANDED HORIZONS

Infinity and infinitesimals are ideas to do with numbers. A rough working definition would be this:

- Infinity *is a number that is bigger than any other number.*
- *An* infinitesimal *is a number that is smaller than any positive number, but is not zero.*

Sound reasonable? Well, maybe ...

Let's suppose that infinity exists, and to annoy all of the mathematical pedants, let's give it the usual symbol ∞. As it's a number, it obeys the ordinary rules of arithmetic, so we can add one to it to get the number $\infty + 1$, which is bigger. That's rather sad, because we said that ∞ is bigger than any other number: now we see that it isn't. So ∞ doesn't actually exist at all.

Similarly: suppose ε is an infinitesimal. Then ε must be smaller than all of the numbers $\{1/2, 1/3, 1/4, \ldots, 1/1066, \ldots\}$. This means that $1/\varepsilon$ is *bigger* than all of the numbers $\{2, 3, 4, \ldots, 1066, \ldots\}$. So $1/\varepsilon = \infty$. But ∞ doesn't exist, so infinitesimals don't either.

These, in brief, are the main reasons why mathematicians were led to reject the notion of infinitesimals and to create, in their stead, Standard Analysis (epsilontics writ large). For historical reasons, they were felt to be very compelling.

However the arguments are really full of legal loopholes that could be exploited. In particular, that weasel-word "number". Just what *is* a number? For example we might choose to take the set of natural numbers $\mathbf{N}$, or the reals $\mathbf{R}$. But they won't do everything we want. Historically, there have been innumerable changes in the meaning of the word "number"; and each has been accompanied by the deafening drumbeat of philosophical warfare as the innovative minority did battle with its reactionary contemporaries. One such step was the acceptance of 0 as a number. Another was rationals. Negative numbers. Reals. Then, with gathering momentum, complex numbers. Then quaternions, octonions ... so dramatic was the deluge that the word "number" was abandoned altogether in the rush to devise new systems of things that could be added, subtracted, multiplied, divided ... and abstract algebra was born. What makes algebra work isn't the objects you use; its the rules of the game you play with them.

The argument above shows that ∞ cannot be a *natural* number. That's not even a surprise: it's obvious. But there's no reason why it shouldn't be some other kind of number. For example, we could invent a whole system of "unnatural" numbers to go with the natural ones. If we wanted the laws of algebra to hold good in the new system, we'd still run into trouble with the following definition:

- *Infinity is an unnatural number that is bigger than any other unnatural number.*

In fact the proof that this won't work would be the same as before. But we'd have no trouble at all if we used:

- *Infinity is an unnatural number that is bigger than any* natural *number.*

The argument works fine up until the point where we introduce $\infty + 1$. This is not a natural number, however, so nothing in our definition forces ∞ to be bigger than it. So at that point the problem with ∞ comes unstuck.

In exactly the same way, it is possible to have infinitesimals, provided they also are "unnatural" numbers.

In essence, this is how nonstandard analysis works. The usual systems of $\mathbb{N}$ and $\mathbb{R}$ of natural and real numbers are enlarged to give $\mathbb{N}^*$ and $\mathbb{R}^*$, the nonstandard naturals and nonstandard reals. Within $\mathbb{N}^*$ and $\mathbb{R}^*$ you can find infinities—whole hosts of them! In fact any number that is bigger than anything in $\mathbb{N}$ is said to be infinite; and this isn't hard to achieve, because $\mathbb{N}$ is a very tiny part of $\mathbb{R}^*$. In fact $\mathbb{R}^*$ can be thought of as a set of equivalence classes (for an equivalence relation defined under the "ultrapower" construction, see below) of sequences (x_n) of reals, where (say) $n \in \mathbb{N}$, and sequences that march off to infinity such as $(1, 2, 3, \ldots)$ or $(1, 4, 9, \ldots)$ then represent infinite reals. There are plenty of such sequences! And if x is infinite, then $1/x$ is infinitesimal, in the sense that it is smaller than any positive non-zero *real* number (in $\mathbb{R}$).

To achieve just this type of extension, however, isn't enough. It's important that the new $\mathbb{N}^*$ and $\mathbb{R}^*$ resemble the old $\mathbb{N}$ and $\mathbb{R}$ to a very high degree. But of course they can't resemble them exactly, or there would be no difference at all. The trick is to decide which features to keep and which to throw away; and then to show that the job can be done at all under those constraints.

Robinson chose with a logician's delicate touch. He insisted that $\mathbb{N}^*$ and $\mathbb{R}^*$ should retain all the features of $\mathbb{N}$ and $\mathbb{R}$ that could be expressed in *first order logic*—which includes all of the usual rules of algebra. For example, since the equation

$$x^2 - y^2 = (x - y)(x + y)$$

holds in $\mathbb{R}$, and since it is first order, it must also hold in $\mathbb{R}^*$. *First order* statements are called "simple sentences" in Hurd and Loeb's book (see p. 11) and are essentially those statements that can be cast in the form $(\forall x)(\forall y)$ [if various properties of the x_i hold then various other properties follow]. Thus, for example, the above statement is $(\forall x)(\forall y)$ [no further properties] $\Rightarrow x^2 - y^2 = (x - y)(x + y)$.

On the other hand, the statement

$$\nexists x \in \mathbb{R} \text{ such that } x \text{ is smaller than any positive } y \in \mathbb{R}$$

(that is, there are no infinitesimals in $\mathbb{R}$) cannot be stated in such a form (in fact it's second order), and $\mathbb{R}^*$ is not required to satisfy it! This is crucial since we want infinitesimals to exist in $\mathbb{R}^*$: that was the whole point of introducing nonstandard numbers.

The possibility of constructing such a system as $\mathbb{R}^*$ is by no means clear; but Robinson showed that it could be done, by using *model theory*, a branch of mathematics that makes a systematic study of when a mathematical system can be found that satisfies a particular set of conditions. By using a construction technique known as an *ultrapower*, it is possible to build $\mathbb{R}^*$ from a very large number of copies of $\mathbb{R}$. Here's a sketch: for details see Hurd and Loeb p. 2.

Let I be any infinite indexing set (say $I = \mathbb{N}$). Define an *ultrafilter* on I to be a non-empty collection U of subsets of I such that

(a) $\varnothing \notin U$.
(b) If A, $B \in U$ then $A \cap B \in U$.
(c) If $B \subset A \in U$ then $B \in U$.
(d) For any subset A of I, either $A \in U$ or $I \setminus A \in U$.

Next consider a cartesian power X of $\mathbf{R}$ indexed by I, that is, a set of "sequences" $(x_i)_{i\in I}$. More formally, consider the set of all functions $x : I \to \mathbf{R}$ and write $x(i) = x_i$. Define an equivalence relation $\sim$ on X in terms of U as follows: $x_i \sim y_i$ if and only if

$$\{i \in I : x_i = y_i\} \in U.$$

Then let $\mathbf{R}^*$ be the set of equivalence classes $X/\sim$. An application of Zorn's lemma proves that non-trivial ultrafilters exist; as an exercise you can now prove that $\mathbf{R}^*$ satisfies all first order properties of $\mathbf{R}$. In particular try the field axioms—that is, for all ultrafilters U, prove that $X/\sim$ is a field.

THE NONSTANDARD APPROACH

By writing out a few calculations, it is possible to appreciate a little of the flavour of Nonstandard Analysis and how it compares with Standard Analysis.

In calculus a typical problem is to find the rate of change of x^2 with respect to x. The usual technique is to imagine a small change from x to $x + \varepsilon$, say. Then the rate of change of x^2 is given by

$$\frac{(x+\varepsilon)^2 - x^2}{(x+\varepsilon) - x} \quad \text{which is} \quad \frac{2\varepsilon x + \varepsilon^2}{\varepsilon}$$

As long as ε is not zero, we may divide out to get

$$2x + \varepsilon$$

If ε is small, this is very close to $2x$. So $2x$ is the "instantaneous" rate of change of x^2 with x.

Not so, said Bishop Berkeley. However small ε may be, it must be non-zero (or else you can't divide by it) so $2x$ is wrong—or at any rate the argument doesn't justify it.

Standard Analysis gets round this by defining the limit of an expression, as ε tends to zero, to be the unique value that the expression becomes indefinitely close to. Since $2x + \varepsilon$ gets as close as you like to $2x$ while keeping ε non-zero, that restores $2x$ as the answer. But at the expense of all that gadgetry with limits.

Nonstandard Analysis tries to have the best of both worlds. If x is a nonstandard real (in $\mathbf{R}^*$) and x is *finite* (which just means it isn't infinite by the definition above) then it can always be written uniquely as

$$x = \bar{x} + \varepsilon$$

where $\bar{x}$ is *standard* and ε is infinitesimal. That is, any finite nonstandard number is infinitely close to an ordinary real number. This real $\bar{x}$ is called the *standard part* of x and is denoted by $\operatorname{std}(x)$. To find the rate of change of x^2 using nonstandard analysis, we let ε be *infinitesimal*, and work out not

$$\frac{(x+\varepsilon)^2 - x^2}{(x+\varepsilon) - x}$$

but just its standard part. This is

$$\operatorname{std}(2x + \varepsilon)$$

which is obviously $2x$ since $2x$ is real if x is, and ε is infinitesimal. Notice that there is no trouble dividing out by ε: although it is infinitesimal, it is non-zero.

Many arguments in Nonstandard Analysis work this way. First extend from **R** to **R***, making available the use of infinitesimals and infinities; then at the end fight your way back into **R** by taking the standard part of the result. It's a startlingly powerful technique, which is very simple to use once it has been set up properly. All the hard work comes in setting it up; but that need only be done once.

APPLICATIONS

Nonstandard Analysis has been used all over mathematics, but seldom extensively, because at first sight it is just a reformulation of Standard Analysis. Indeed, a particular result is either true in both, or false in both: Nonstandard Analysis is not capable of proving any theorem that cannot be given a Standard proof. At first sight this translates as "Nonstandard Analysis is no use"; but that takes no account of the relative simplicity and efficiency of the new technique. In principle nothing can be achieved by a bulldozer that could not be achieved by men with spades: that's why nobody in his right mind ever uses a bulldozer, right?

Just so. And the extra power of the Nonstandard approach is beginning to be felt, in particular in certain problems in applied mathematics, especially in *perturbation theory*, which studies the effects of *small* variations on a problem. It is a simple step to study infinitesimal variations, using nonstandard analysis; and because it is a logically rigorous technique, it can be pushed to its limits without fear of error.

Consider, for example, the oscillations of an old-fashioned thermionic valve, as used in radio sets a little after the time of Noah (that is, up to the 1950s). The onset of such oscillations is defined by a system of equations, and these may be studied by drawing a "phase portrait" in which closed loops correspond to oscillatory solutions.

When the oscillations switch on, they do so by "growing" a tiny loop from what was originally a steady state. At some later stage this loop has become very large. The question is, how exactly does it grow?

Until recently, everyone assumed it just grew; that is, nothing very interesting occurred *as* it grew. But by using Nonstandard Analysis it became clear that a very remarkable process happens. First the steady state produces its tiny loop, as was known already. but then, very suddenly indeed, the loop expands by a massive amount. In the middle of this rapid but in principle continuous change, the loop acquires a shape that looks rather like a duck. Hence the current name for the phenomenon, "canard", from the French. (Papers on the subject in professional journals include "The Duck Hunt" and "Nessie and the Ducks": for some reason, activists in Nonstandard Analysis seem to have a more evident sense of humour than is common in the mathematical literature.)

Another application of current interest is in the billiard problem. Imagine a billiard table whose rim, unlike the usual one, is an arbitrary closed curve. Let a billiard ball bounce indefinitely off this rim. What will it do? It's actually a very important problem in dynamics, because it is one of the simplest stages in a study of how molecules in a gas bounce off one another and off the walls of their container: here there's only one molecule. Nonstandard Analysis has considerably simplified the proof of several basic theorems in mathematical billiards.

A third example of the power of Nonstandard Analysis is its use in the theory of stochastic differential equations. This is a highly technical but important area of mathematics with applications throughout science. It deals with the motion of a system that obeys a differential equation, but is subject to random disturbances. Think of playing cricket in a gusty wind: what does the ball do? A heuristic description of a stochastic dynamical system is that at each infinitesimal instant dt of time it moves an infinitesimal distance $dx + dw$. Here dx is prescribed by the rule $dx = (dx/dt)dt$, and dx/dt is determined by the underlying differential equation $dx/dt = f(x, t)$. The second term dw is an infinitesimal random disturbance, having some specified probability distribution.

In classical analysis this description is nonsensical. The usual technical definition of a stochastic differential equation is therefore far more complicated and indirect: instead of having a single solution it has a space of solutions with a probability distribution, and so on. However, Nonstandard Analysis allows us to implement the above description *rigorously*. The result is a major simplification of the technical apparatus used in stochastic differential equations.

Other applications include boundary-layer flow in fluid dynamics, solitons, diffusion and chemistry. It seems likely that as the methods become more widely known, more and more pieces of mathematics that involve a "small" parameter will turn out to be comprehensible from the point of view of an infinitesimal parameter. If so, there ought to be a great deal of unification and simplification of this difficult and technical area of applied mathematics. More to the point, after enough experience has been obtained, the more powerful technique may be expected to yield new results that are just too complicated for the more traditional methods. It will happen only against resistance from some of those who are well-versed in the traditional methods and hence see no reason to change, and it may not happen at all, but there's a quiet revolution brewing. Who knows: by the twenty-first century, maybe all engineers will have to be trained mathematical logicians as well.

FURTHER READING

Deledicq, A. and M. Diener, *Leçons de Calcul Infinitésimal*, Colin, Paris 1989.

Diener, M. and C. Lobry, *Analyse non Standard et Représentation du Réel*, OPU, Algeria 1985.

Hurd, A. E., *Nonstandard Analysis—recent developments*, Springer, Berlin 1983.

Hurd, A. E. and P. A. Loeb, *An Introduction to Nonstandard Real Analysis*, Academic Press, New York 1985.

Keisler, H. J., *Foundations of Infinitesimal Calculus*, Prindle, Weber, and Schmidt, Boston 1976.

Lutz, R. and M. Goze, *Nonstandard Analysis—a practical guide with applications*, Springer, Berlin 1981.

Stewart, Ian, *The Problems of Mathematics*, Oxford University Press 1987.

What Happens when an Engineer Learns some Maths

Or, On Archimedeans: A 16th-Century Tragi-Comedy

Piers Bursill-Hall

As every schoolchild knows, the Scientific Revolution—when heroic modern science was born and the Dark Age of a geocentric world view, and a methodology that centred on philosophical speculation and neglected experimentation, was overthrown—occurred sometime in the seventeenth century, around the figures of Galileo and our own Local Boy Made Good, Newton.

THE "GREAT MEN" OF THE SCIENTIFIC REVOLUTION

In the official Pantheon of Heroes of Science there are not a few who are recalled by the folk-histories as being Heroes of the Scientific Revolution. Working between the last years of the sixteenth century and the late 1610s, Kepler, of course, finally got astronomy "right" by discovering the orbits of the planets to be ellipses. It is convenient to forget that he made this discovery for all the "wrong" reasons: his motivations were far more mystical and theological than we would care to admit in a "modern" scientist, and his calculations haphazard and directed by a density of calculational errors that beggar the imagination; and that Kepler's "three laws" are only three that have subsequently been selected as "correct" out of the dozens of proportions and relations that he found that he thought had theological, harmonic or physical significance.

In the late 'teens to the '20s (though not published until the '30s and '40s) Descartes established the link between algebra and geometry—the single most important event in the millenium, I would suppose. He also began the rethinking of natural philosophy and the foundations of physical explanations by proposing that all macroscopic events could be explained by the mechanical activity of microscopic corpuscles. His laws of corpuscle motion are almost all wrong (some trivially so), and the promise of a mechanical reduction of phenomena turned out to be pie in the sky ... but it was more than appealing to contemporaries and successors.

Newton, of course, hardly needs mentioning; we have all taken our parents/Great Aunt Gladys/pet dog around Trinity Chapel and explained that Newton in the 1660s to '80s invented the calculus, explained the motions of the heavens, uncovered something about the nature of colours with his prism (although quite what, we fudge quickly), and more or less solved all the problems of the world except poverty, war, pestilence and inflation.

But in this Pantheon, surely there is no hero figure so heroic as Galileo. Working in Italy between roughly the 1600s and 1630s, Galileo was the inventor of the Experimental Method and martyred by the (nasty, backwards) Roman church for his belief in a sun-centred universe. To Galileo we owe the invention of the telescope (never mind if some Dutchman had done it before), the discovery of some of Jupiter's moons, the successful analysis of free fall, some sort of experimental and mathematical analysis of inertia and projectile motion in general, and the experimentally based development of a mechanics that demonstrated the heliocentric model of the universe. All these combine to create the image of Galileo as the virtually single-handed founder of modern physics and its methodology. Add the powerful drama of an old man forced under torture by the dusty clerics of the Inquisition to recant what he knew to be truth—and the folk-history of science has a potent figure, saint and martyr. Not for nothing are institutions, satellites, celestial objects and the like named after him. To find himself portrayed as such an heroic scientist would only have pleased Galileo: as a master publicist, manipulator and media figure, this is something like the image he would have liked to create.

1. A Problem with the Scientific Revolution: Copernicus

But there is an interesting problem. If we look *back* just a little, we might notice Copernicus—almost forgotten in this folk-history—whose heliocentric model of the universe was (so the story goes) the crucial step in breaking with past science. One might notice that Copernicus had the simple notion of a sun-centred universe sometime around 1500 (between 1497 and 1510) ... and that his argument—and detailed calculations—for the heliocentric model were published in his *De Revolutionibus* of 1543. Notice the time gap: Galileo published his telescopic evidence for a Copernican model of the heavens in 1610; but his historically much more significant justification of the Copernican model on grounds of experiment and physics had to await his *Dialogue on the Two Chief World Systems* of 1632.

Copernicus, of course, didn't invent the idea of a heliocentric universe out of thin air—it came out of a context that enabled him to be interested in the problem, and to see heliocentricity as a plausible solution to the problem of planetary orbits. This was around the year 1500, and the publication was in 1543 ... and then there is some *seventy years* during which Copernicus' model is largely ignored; and ninety years after the publication of *De Revolutionibus* Galileo publishes an attack on the traditional Aristotelian-Ptolemaic system and a defence of this Copernican model, and gets into severe trouble with the Church.

One might get suspicious. If Copernicus' heliocentric model and the reform of astronomical thinking has something to do with the Scientific Revolution—and all the signs are that this was the case—then what is the reason for this seventy or ninety year gap? Why did Galileo not get into trouble with the publication of the *telescopic* evidence for Copernicanism in 1610? Evidence which, to our modern eyes, is much more convincing than the 1632 *Dialogue*. Why does the Church not get upset about Kepler's various Copernican publications from the late '90s to the late 'teens? And why this 70 to 90 year gap between the publication of *De revolutionibus* and the reaction? Ignoring the telescopic evidence, on which Galileo does not rely in the 1632 *Dialogue*, one might think there was little in the *Dialogue* that couldn't have been done in 1550 or so: what is the problem?

2. Galileo and the New Science

You should sense that something was going on in the intervening years that meant that Galileo was doing something different in the 1632 *Dialogue*, or that his project was different from what one might think up in the context of 1500 or 1543. And you ought to suspect that Copernicanism *tout court* was not the only central issue.

First of all, let me remove Galileo's undeserved titles. A saint he certainly was not. A thoroughly unpleasant and argumentative man, he managed pretty systematically to alienate most of his colleagues when he was Professor of Mathematics at the University of Padua; still more so his philosopher [read "scientist" for this term in the 15th–18th centuries] colleagues when (after 1610) he was Court Mathematician and Philosopher to the Grand Duke of Tuscany. Not so much because he disagreed with them, but because he couldn't resist disagreeing with them in the most violent, vicious and personal terms, which in Britain today would have had him up in court for libel most weeks. He never did marry his wife, and when it was financially more convenient, he packed his daughters off to a convent ... Not exactly a rôle model for the budding young scientist today. He was no martyr to anything other than his own ill temper and determination for self-promotion.

If he had published his later works defending Copernicanism in Latin instead of in racy, readable, and very, very popular Italian, nobody would have much bothered that he was disobeying the Church's edict not to publish Copernican works. Had he not tried to write to, appeal to, and publicise his book widely to an educated lay audience, but addressed it more carefully to technical astronomers and philosophers, nobody would have cared overly. Had he just gone though a soon-to-be-common ritual of saying that the Copernican view was just an *hypothesis* instead of asserting that it was absolutely, unquestionably certain, and physically true of the universe, Rome probably wouldn't have read past the Preface. Galileo's problem was that he couldn't let well enough alone, and tried to promote himself into the limelight and to meddle in Church politics ... and eventually got his fingers burned. He became a convenient way for some Vatican Curia politicos to annoy his patron, the Grand Duke, and to remind the dissident chattering classes to be a bit more careful about following the Party line. And his martyrdom? His trial was pretty much a bored formal affair, and the Church's lack of interest in him is very clear: all they wanted him to do was shut up and they would be happy. Galileo wasn't even shown the instruments of torture (a ritual required for the recantation of his errors to be valid), and they didn't confiscate his (considerable) wealth; he was put under house arrest for a few months with a Bishop friend of his, in the Bishop's exquisite palace (hardship, hardship), and then sent back to his beautiful country house outside Florence, where he was left unfettered, and the Church politicians got back to more important matters than some nitwit dissident philosopher. Hardly anything more than the most ritualistic rapping on the knuckles.

3. Galileo and the Experimental Method

But if he wasn't a martyr, at least he *was* Saint and Father of the Experimental Method, surely? Again, not really. In the *Dialogue* he reports a lot of experiments that build up the physics he needs to justify a Copernican model of a moving Earth, and if you read it carelessly, you might even think he was reporting experiments that he had done. But no, he had not in general done the experiments, and indeed doesn't purport to have done them. Most of them are just thought experiments, or appeals to carefully thought-through everyday experiences, or are very clever arguments that propose an experiment and then go: "but we don't need to do the experiment, as I can make the result clear to you through reason and logic, bringing you to knowledge of the answer without appeal to your senses"—obviously much more persuasive than having to bother to go and actually do a boring experiment. Not only did Galileo not do many of the experiments reported, but he turns out (in the argument of the *Dialogue*, at least) not to have regarded experimentation as all that sure a method of obtaining knowledge of the world: he preferred reasoning from common experience or *a priori* from first principles, not empirically. Exactly the methodology he was supposed to be overthrowing. Oh, dear.

However, two points should be quickly made—points that will eventually be explained a little further. (1) By reading carefully the odds and ends of papers that Galileo left behind, scholars have determined that he was, in fact, doing experiments—and brilliant, careful, insightful investigative and analytical experiments—throughout his life. What he was doing back home in the lab, however, was not what he reported in the *Dialogue* ... but it turns out that he was an absolutely superb experimentalist. (2) Why, then, does he not report these experiments in his published work? Why refrain from using them as support for his arguments? Because Galileo does not think they would convince his audience: he does not think that the experimental methodology would be considered a convincing one. And indeed he was right. The standard educated layman of the day would not have regarded a methodology so flawed and so fraught with difficulties as experimentation (remember *your* sixth-form labs and all the fudging you had to do to get the "correct" results?) as a reasonable way to obtaining sure and certain knowledge. It was much more convincing to argue that sure and certain results—results of the status of science—were to be had by pure logic ... which is a line of reasoning not at all strange to the modern day mathematician, after all ...

4. Experimentation as a methodology

In any case, doing experiments was hardly new in 1630: the idea of doing experiments to sort out your problems was commonplace, and to be found in Europe as far back as the twelfth or thirteenth century. So who might have been "the Father of Experimental Methodology"? Sadly there is none. But we cannot simply date the beginnings of the experimental methodology in the middle ages, a time when we know that empirical methodology was widespread in certain circles—nor is life so simple that we can find the origin of the Scientific Revolution in the 1200s, just because certain folk were proceeding to solve their problems using empirical ways. Simply *doing* experiments is no particularly big thing: what matters is what you think their status is, what sorts of truths they reveal ... whether you think that they are the way to find out the scientific truths of the nature of the world.

Experimentation as a methodology goes back as far as you care to look. For example, we have second century AD Hellenistic texts on what might be called chemical experiments, and the great engineering and technological feats of the Greeks (let alone the Romans) were certainly not achieved without a great deal of perfectly systematic experimental study. Ancient Greek doctors and medical thinkers were empiricists from as early as we have records, and dissection, for example, was part of their way of studying anatomy and physiology from the fifth century BC. Consider the great fifth century Greek oared battleship, the trireme, which had three rows of oarsmen packed closely together to produce what was a remarkable fighting vessel. On the basis of a few fragmentary descriptions that have come down to us a trireme was recently reconstructed by a Cambridge based team and a replica built. Getting the three rows of oarsmen together in the required space so that they could still row even near to optimally proved to be an extremely difficult engineering design problem. The team had recourse to the best modern naval design skills and techniques, and it still took no small effort to get the design to work practically and optimally. Naturally, it turns out that the descriptions of the trireme that we have do fit well with the optimal design that was found: the Greeks were not the origins of Western civilisation for nothing. Lacking computer simulations and the like, it is clear that they could only have designed such an optimal configuration of oarsmen—as well as things like clever means of making the ship strong enough to support its own weight—by means of a great deal of careful experimentation.

If there were experimenters amongst the classical Greeks and Romans, all the more so by the Western Middle Ages. The Mediaevals loved to make things bigger and better (think of the size of mediaeval churches and cathedrals), and they loved gadgets, widgets, mechanical devices and labour-saving machines. Warfare, agriculture, and manufacturing all developed remarkably after the "re-awakening" of the middle ages in (say) the later eleventh century, and their technologically oriented, machine-minded development all speak of an empirical, experimental approach that was both widespread and clearly very successful. In the late Middle Ages and earlier Renaissance (the late 14th, 15th and early 16th centuries), this fascination with technology grew, engineers achieved a new sort of social stature, and innovations in the technology of manufacturing and warfare became some of the central pressures in the economic and political developments in the civilised world. New chemical, metallurgical and manufacturing technologies, along with radical improvements in the technology and "arts" of warfare were simply to transform society ... and places like some of the city states of Northern Italy—Florence is the most famous—were to become enormously rich and powerful, on account of these technological innovations based on an empirical, experimental approach to solving problems.

In the sixteenth century this was to become a stronger and stronger phenomenon. Take the empirical study of alchemy—a pejorative term to us, but in fact simply chemistry being done before the more-or-less-modern causal mechanism of chemistry as particles combining and separating came to be accepted. Alchemy became one of the hot new scientific studies: trendy, radical, and with possible industrial applications. Alchemical studies were closely linked to the metallurgical, tanning and dyeing industries, and in the hands of Paracelsus and his followers were to promise vast and wonderful new improvements in medical drug therapies ... some of which even worked. The promise of better medicines and better medical care was something that, none too strangely, seemed quite important to the people at the time.

So, to cut this long story short, by the end of the sixteenth century experimentation was positively commonplace in Europe. Between civil and military engineers and such people as the alchemists and the alchemical pharmacologists, doing experiments was hardly innovative: it was virtually a standard way of approaching and solving large classes of problems. But classes of problems that were not regarded by the thinkers of the day as what we would call *scientific* problems: there's the rub.

5. Galileo: No Saint, Martyr or Father

So much for Galileo as the inventor of the Experimental Method. Yet Galileo and his friends and followers must have been doing *something* different, unless one is willing to believe that the history is simply a complete fantasy. If we agree that he was no saint or martyr, and not the inventor of the experimental method as such, you might be driven to conclude that he did nothing out of the ordinary at all, except in that he was a media figure who managed to become a seventeenth century television personality (so to speak). Well, there is (of course) some substance to the historical image, and there is great substance to the idea that we see in Galileo something of great importance, which precipitated a profound changing and development of ideas in science. It is not so much to do with doing experiments *per se* as it is to do with what you believe experiments tell you, what the study of nature by the use of experiments tells you ... and above all, with the notion (the incredible, *completely* unintuitive notion) that the experimental study of mechanical phenomena gives rise to some sort of paradigm, sure and certain mathematical science, to which all other sciences should approximate as closely as possible.

This is an almost absurd notion: most of the phenomena of the world that we experience are simply not particularly mechanical: growth, change, life, death, and everything—admit it: the world around you is simply not very mechanical, and if you had not already been taught it (read: been brain-washed into the idea), it would seem patently absurd to suggest that this very organic, chaotic world was in some way underpinned by some sort of mechanical notions.

This is, in a rather inexplicit way, what Galileo was to suggest; and it is for this that his work is so remarkable. He was able to suggest and argue, reasonably convincingly to at least some of his contemporaries, that in some sense mathematical mechanics is a paradigm science, that mechanical notions allow us to subsume a wide variety of phenomena to mathematical analysis, and these phenomena ultimately act mechanically ...

What is interesting—just for the moment—is to consider how this might have come about. For it is not clear (in this folk-history of science) that there was a tradition before Galileo that furnished him with these notions: (1) that mathematical mechanics was some sort of unifying, paradigmatic and fundamental science, which would allow him to unify the causal mechanisms of both celestial and terrestrial phenomena, and (2) the idea that experimentation on mechanical phenomena might lead him to a successful analysis of such diverse things as local free fall and the motion of the planets. Why look at experiments as leading to some sort of certain knowledge of the world, and why think that mechanics might be some sort of sure and certain general science of physical phenomena?

The answer, a horrifying answer for any modern day mathematician or theoretical physicist, lies with what happened in the sixteenth century as a result of a series of otherwise wholly unexpected circumstances amongst a community of (albeit rather well-educated) engineers: they learned some mathematics, and some of the natural philosophical (read: scientific) implications of such mathematics ... and as a result they started to break all sorts of rules, and all hell broke loose. You might note that in general ever since then engineers have been strictly forbidden to do any mathematics or proper science for fear that they might upset the comfortable apple-cart of science again.

RENAISSANCE ENGINEERS

The particular conflation of engineering interests and mathematical capacities and ambitions could only have happened when it did—in the 16th century—and only as a result of the quite extraordinary circumstances of the time.

Cast your mind back (so to speak) to the thirteenth century. Europe was deep in the *soi-disant* Middle Ages caused, at least as far as the history of things scientific need be concerned, by the general intellectual (scientific, to be a little less over-generalised) degeneracy and then collapse of the Roman Empire in the Latin West.

1. Transmission of Ancient Texts to the West

The Romans—the Romans of the Western Empire, to be precise—were hardly a very scientific people. They were proficient at such things as empire building, corruption, warfare, and that sort of thing, but simple things like elementary mathematics and science were rather beyond them. This shouldn't surprise you: consider multiplying 276 by 819 ... in Roman numerals. Adding and subtraction are more than complicated in the Roman system; serious multiplication and long division are just out of the question. They would have to be done by professional calculators on counting boards or abacuses. Indeed, such was the level of technical ignorance within the Western Roman Empire that if you wanted to know anything or learn anything natural-philosophical or scientific, you hired a Greek tutor and learned it *in Greek*. Some of the most senior administrators in the West were Greeks, and we have examples of surviving lists of the personnel involved in some large engineering projects—and the key positions that required mathematical or technical competence (beyond the artisan's rule of thumb) such as the engineers and surveyors, were Greek names. The reason is perfectly simple: the Greeks were the better educated, the mathematically and technically more able people of the Empire, and so naturally they were drafted into positions that required such learning. The Romans themselves seem only rather more rarely to have managed much of a mastery of Greek science and mathematics—and when they did, it was generally only to remain derivative of the Greeks.

With the collapse of the Western Empire, what little learning there was in what is now Western Europe pretty much disappeared, leaving everyone in the Dark. Remember, however, this was only a very *local* phenomenon: by the last few centuries of the Roman Empire, the Greek end of the Mediterranean had been divided off into the Eastern Roman Empire, capital at Constantinople; the Eastern Empire may have had its difficulties with inflation and unemployment and the exchange rate mechanism and what not, but they were to carry on having basically a good time through the *Western* so-called Dark Ages. When the Mediaevals decided to be Mediaeval, and generally stop being so thoroughly Dark, they had little scientific, technical or mathematical knowledge of their own, or easily and locally available to them, upon which to build.

Rather they found themselves turning to the Arabs, obtaining from them first rather pedestrian texts in the sorts of technical issues that interested the (now tenth and eleventh century) Mediaevals, and then discovering that the Arabs had to hand, amongst other things, all sorts of learning that they had obtained by rather indirect routes from the classical Greeks—translations of Greek texts, Arab compilations, teaching texts and commentaries, and the like. In general the original Greek texts of the ancients were not available to the Mediaevals, of course: what they were able to get from the Arabs was in Arabic, and had to be translated into Latin.

This was not a trivial task: imagine the difficulty! You want to translate a technical and subtle subject like Euclid or Archimedes into Latin, but of course since your (eleventh century) Latin has no strong pure mathematical tradition behind it, it doesn't have the technical terms ... indeed, even a competent translator would have no idea what the text was talking about, simply because such mathematics was vastly above the level of anything the early Mediaevals would have ever seen. So the language literally doesn't have the technical terms to translate into ... and you have to translate the text coherently, and then learn and appreciate its contents so that you can understand what the mathematics is about ... before you can translate the text. Add to this the problem that the texts the mediaeval scholar would obtain would often be thoroughly corrupt—from several translations into languages that didn't have a technical capacity to "receive" the translation, from scribal errors (replacing a "+" with a "−" ... or much worse), missing proofs, diagrams, words, lines, chapters, pages ... all the sorts of things that can happen to a handwritten text that only survives by being selectively hand-copied over the centuries. The problem of translating, making sense of, interpreting, and understanding the classical Greek inheritance that came to the Mediaevals via the Arabs ought to seem to the modern observer to be a virtually impossible task.

Remarkable though the mediaeval achievement was, it was a very problematic and incomplete reconstruction of the works of the Ancients. The reasons for this were not far to seek: they lacked many of the ancient texts; many of the texts they had were corrupted or obscure because of the transmission process; and, perfectly naturally, the Mediaevals had learned their ancient science, natural philosophy, geometry, and what have you, through the mediation of the Arab teaching texts and commentators. They simply could not take even a good translation of Aristotle or Euclid and teach it to themselves without any preparation: they had to learn these texts through the teaching and explanation—and therefore the eyes and interests—of the Arabs. Towards the end of the fourteenth century, some scholars became increasingly aware of and dissatisfied with this situation, and began to cast about more seriously than ever before for better texts and for those other texts that they were told in the then extant sources had existed, but they no longer possessed.

It was not long before some scholars discovered a gold mine right next door in Byzantium. The Byzantines—just the Eastern Roman Empire a thousand years later—were still *Greek*, and had maintained classical Greek as their official theological and scholarly language. The Byzantines had also been decent enough to maintain many of the Greek and Latin manuscript texts that had survived into the seventh and eighth centuries AD, so in, say, 1400, a Western scholar could find a vast array of classical Greek and Latin texts in Byzantine libraries and monasteries ... texts that had suffered much less corruption and degradation due to translations and the filter of selection and abridgement, and were free from the additions and commentaries of the Arabs. It was a mediaeval scholar's paradise.

Western scholars, primarily Italians (Italy being the Number One nation at the time, and, via the Church and via Venetian and other seafaring traders, having reasonably good relations with Byzantium) began to travel to Constantinople to buy or copy these manuscripts. Their interests were not primarily scientific or technical of course: they were interested in pretty much anything classical. This wave of Byzantine texts began in the last decades of the fourteenth century and manuscripts continued to be brought West until the Byzantine empire finally succumbed to Muslim invaders in the 1440s, ending the last direct links to the ancient world, and turning off at the main this flow of manuscripts. And thus it has remained: there is very little ancient learning that we have subsequently recovered (there are a few spectacular exceptions, like Archimedes' Method, rediscovered in 1906—but that is another story).

This *second* wave of ancient texts, this Byzantine inheritance, was to change the (Western) world. These manuscript collectors and library builders were not your impoverished, down-trodden and oppressed scholars of today, but gentlemen of learning, wealth, social status and position. Early in the fourteenth century the "*grecisti*" (as they called themselves in Florence) hired Greek scholars to teach them (and to lecture publicly, and to hold newly founded University Chairs) in Greek language, literature, and learning in general. The fashion of the study of these Greek texts, and the learning they held, soon spread to the (Italian, and later other) Universities, and to their faculties of Arts (NB: the faculty of Arts then includes what we would call sciences, and were opposed to the other faculties of Law, Medicine and Theology). The "arts" or "humanities" students and professors thus enthused with these new studies, and then the scholars, learned gentlemen, Patrons and nobles, came to be called "humanists" (from the slang term for students in the arts faculty; nothing to do with the modern meaning of "humanism"). The humanist movement and influences were to form, in some sense, the intellectual backbone of the Renaissance, and it is the extraordinary flourishing of everything—fine arts, architecture, literature, science, commerce, technology, whatever—in the Renaissance that was so completely to transform the Western world. And, in no small part, it stems for the effects of this influx of manuscripts from the East, and the re-awakening, re-birth (yes, re-*naissance*) of the glories of ancient learning, culture and civilisation that humanist intellectuals and patrons of the time thought they could at last bring about.

2. Renaissance Humanism

The earlier humanist scholars were, quite naturally, much more interested in the new, improved texts and translations of works they already knew about—Aristotle's philosophical and scientific works, for example. But, as the old saying doesn't go, "with the baby you get the bath-water". Not only were there new and better versions of the texts the Mediaevals already had, but there were many others—the existence of many of which the Mediaevals had had no inkling. Of the most obviously cosmic importance was the recovery, translation, reading and interpretation of Plato's works, and the late Hellenistic followers and commentators of Plato—like Proclus, Pappus, and others.

For reasons that would take us into an intolerably long aside, the reading of Plato and the "neo-Platonists" was to have a stunningly powerful impact on later fifteenth and sixteenth century philosophical and natural philosophical thinking. Plato was much admired by the Byzantine scholars (who were properly much lionised by the Italian humanists), and had inherently great authority because Plato, pupil of Socrates, was the teacher of Aristotle ... and was clearly the pivotal thinker in the Greek development of philosophy.

Even more important was the apparently proto-Christian nature of Platonic and neo-Platonic philosophy. Interpreting and understanding Aristotle within a Christian context had not been all that easy: St Thomas Aquinas did this in the thirteenth century and ever since everyone has recognised this as the single most significant intellectual feat of the Middle Ages. But Plato, and particularly the Hellenistic neo-Platonic commentators (who are not Christians), sound just so Christian it is eerie. Some humanist neo-Platonist enthusiasts argued that Plato represents a sort of pre-Biblical revelation ... God setting out a certain amount of the Word in a pre-Christian philosophy before laying it out once and for all, fully but in a different form, in the Bible. This, in a Christian world like the 15th and 16th centuries, was to give Platonism a quite fantastic authority. It was not a difficult idea to accept at the time; the Old Testament was, after all, an accepted pre-Christian prophecy or revelation. If you read Plato's *Timaeus* you will see that the interpretation of *Genesis* that we have (and which early Church Fathers worked out) owes more in some aspects to the *Timaeus* that it does to *Genesis*. This is, in fact, hardly very surprising: the early Church thinkers who interpreted scripture and the Christian revelation, working out Christian theology, rites and rituals, laws, and doctrines, were working in a Hellenistic world heavily influenced by neo-Platonism. They themselves were profoundly influenced by it, and expressed much Christian theology in a philosophical language that is deeply neo-Platonist. St Augustine is a prime example of this. Indeed, Kepler, at the beginning of the seventeenth century, was perfectly clear on the matter: for him, Plato's *Timaeus* was of Biblical status, and its revelations (which Kepler had found in it, of course) were to be treated accordingly.

3. Engineering Texts and the Renaissance

However, the humanists were not only concerned with reading Plato and the neo-Platonists: such was the authority that these classical texts commanded, such was the respect that humanist scholars had for them—a respect only encouraged and sanctioned by the stature given to Platonic thought—that *any* classical text was studied with enthusiasm and reverence. And here follows the wrinkle (or the bath-water): because of the respect that the humanists had for all things of the ancient world, it followed that these were the proper domain of study, or interest, or enthusiasm for socially upper-stratum scholars, Gentlemen, and the socially high-status, learned patron. The consequence is a minor, sly, tiny little crack in the door that was to usher in a small change of almost no consequence. The reverence that the humanists had for the ancients and their science, art, architecture, or whatever, meant that *whatever* they had of the Greeks was noble and worthy of study by the scholar and (enlightened, learned, high status) Gentleman or Noble.

This is what will allow some things to change. Had the humanists found an ancient Greek treatise on—say—hairdressing (and why not? A long letter from the chief of the Athenian Guild of hairdressers and haircutters, addressed to a young apprentice on how to cut and dress hair ... it is almost imaginable), then the translation, study, and commentary upon such a text would have been a fit and proper thing for a humanist to do. And there would have ensued further treatises, commentaries, and learned discussions on "hairdressing for today, based upon the most noble and ancient principles" and the Noble Science of Hairdressing, Fit and Proper for Gentlemen and the Young Nobleman. Not, I hasten to add, that in this fantasy case the scholar, gentleman or noble would be encouraged to go and set up shop and engage in practical hairdressing: obviously this was a socially much lower level craft practice, hardly the domain of the Gentleman ... and the artisan's craft practice was, and would remain, quite untouched by the Gentlemanly humanist study of the Greek text.

The serious point, however, is one of the humanist respect and enthusiasm for the classical Greeks actually working to permit changing social and intellectual positions of subject disciplines. Something that was not proper for a Gentleman to engage in could, in a suitably non-practical, non-menial form, become something within the proper domain of the scholar and Gentleman's study if it had the sanction of some sort of classical Greek precedent. And, leaving aside hairdressing, this is what was to happen in the case of certain aspects of what we would today call engineering. Because amongst the detritus (the bath-water) of the texts which came from Byzantium, and were so highly valued by humanist collectors and scholars, were the texts of Vitruvius, Hero of Alexandria, Pappus, Philo, the Pseudo-Aristotelian "Questions on Mechanics" [thought at the time to be by Aristotle, but by the seventeenth century recognised to have been written by a follower in the third century BC], and others ... which were texts on more or less theoretical aspects of machines, mechanics, or more practical applications and engineering.

It is important to recognise that this material, these "engineering" texts (permit me to use this anachronistic term just for ease) occupy a problematic and ambiguous place in the intellectual and social categories of the time. These are texts of the ancients, and they show an understanding and a perception that was something of a revelation to the thinkers of the time. To any self-respecting and technically lightly-literate late 15th or 16th century humanist, they are texts of the greatest interest and authority: they are the proper object of study of a scholar. But they are concerned with the practical, mechanical arts, the domain of the artisan, the craftsman, the "crude mechanic" (to use a typical seventeenth century turn of phrase). The contradiction is easy enough to explain, because these texts give us the theory, the "philosopher's study" of these arts: the works of Hero or Pappus are not about what size drill to use here or how to mix mortar ... they are about the mathematical description of simple and compound machines.

4. Engineering and *Scientia*

Now the philosophers-cum-scientists are set loose on this stuff. What is the status of this more theoretical study of the principles of the practice of the mechanic or artisan? Is this an Art, or is it a Science? The distinction is important, because a Gentleman does not engage in the practical Art—that is what menials do—but a Gentleman may engage in a Science, since this is cerebral, speculative, philosophical. The different social siting of the disciplines is due to their different epistemological status. In the categories of the time a science (***scientia***, meaning true knowledge and understanding) is part of the pursuit of philosophy, "to have knowledge"; and art (*artes*, meaning crafts or skills) is about the practice of doing things that are not natural, or are opposed to the action of nature. The distinction is between the necessary and timeless truths, which are the subject matter of the sciences, and the contingent truths of the practitioner or the art.

This distinction between necessary truths and contingent truths is important, and worth understanding clearly. The truths of science were supposed to be absolute and necessary, timelessly correct. Now, over the last century, some scientists have changed their tune: no longer is science about the Truth, or what is absolutely right. It is about explanations that are more or less adequate, that seem to work or at any rate haven't yet been proven wrong. But this is a recent change in the epistemological status of scientific knowledge: since the early Greeks, the idea had been that science was about sure and certain, timeless and necessary truths, and at any rate certainly not the contingent truths of happenstance.

To get a feel for the distinction, consider the following two statements: (1) there is *not* a lion (fully grown, standard issue) in the room where you are reading this, and (2) Socrates, Athenian citizen and teacher of Plato, was a mortal. It is likely that you will have no trouble accepting both these statements as true (barring legalistic quibbles). There is certainly not a fully grown lion in the room I am in right now—so (1) is true for me, at least (and I'll bet your Poll Tax bill that it is true for you, too)—and Socrates was obviously a human, and hence had one of the properties of being-a-human, which is mortality—so (2) is true. But consider the quality of the truth of (1) and (2): whilst (1) is true, it is contingently true because it *could* be otherwise: unlikely, but you could be reading this in a room with ... and so on. But *could* Socrates, a human, be anything other than mortal? No: it is one of the properties of humans to be mortal, so this is not contingent upon the happenstance of the world, it is a necessary truth.

Science in the 15th–17th centuries was supposed to be about truths of the second kind—necessary truths: the truths, the rules and the knowledge of the artisan or practitioner may well be perfectly true in the sense that what he says will happen *does* happen, but they are not necessarily so, only contingently so.

So what status should be given to the truths of the theory of mechanical phenomena that the enthusiastic humanist scholar or teacher would teach his enlightened humanist gentlemen students or noble patrons? The answer required a little care. For the objects of this study—say the simple or compound machines, or simple principles of architecture—were clearly physical, changing, and artificial, so knowledge of them would be contingent; yet the Greek treatises had all the nobility, and social and intellectual cachet, of a science.

The interest in these classical "engineering" treatises was quite widespread in the 15th and 16th centuries, and we find them being taught quite regularly at the fashionable, humanist-influenced Universities and "Gentlemen's Academies" in Northern and central Italy by the 1520s and '30s. These Academies, for "Gentlemen of Letters or Arms", would teach the sons of the nobility and near-nobility things like the principles of engineering and architecture. Not that these lads were going to go and dig canals, build houses or fortresses, make useful machines for lifting weights and so on *themselves*. Rather, as landowners, rulers, or military commanders they would be in a position to oversee and to patronise such projects—and thus better educated, they would be better fitted to do so. This, in context, was perfectly reasonable.

But the consequence was, in effect, to begin to offer the possibility of raising the social—and intellectual—status of the practitioner, the "engineer". From early in the sixteenth century we find examples of—effectively—engineers who try to raise their own status from "menial" to "gentleman" by taking on the humanist trappings of scholarship and learning. It is one thing to say to a patron "I propose to build fortifications for you thus-and-thus", and another to say "I shall design you fortifications on the noble and timeless principles of the great [whoever, as appropriate], using the best and most ingenious machines that I have developed from my deep and profound study of the manuscripts of [someone suitable and recondite] and the principles of machines I have found were discovered and explained by [another suitable classical authority], so that you, most illustrious and noble lord, in whom the true light of science and learning of [say, Alexander the Great, just to be modest] shines anew, shall be like [yet another classical hero, on some suitable, noble/victorious place/occasion] ..." and so on and so forth, laying it on with a trowel. You get the picture.

Clearly the engineer in the latter case is both asserting the dignity (and the humanist-legitimised status) of his work, and his own status as a man of learning and someone quite distinct from the mere artificer and practitioner. And we find plenty of examples of this sort of humanist-levered upward mobility on the part of 16th century engineers (there is a joke about early yuppies somewhere here).

5. Engineering and Archimedes

And then, around the middle decades of the century, these engineers fell upon the most extraordinary tool: Archimedes.

Archimedes' works had, in fact, been known to some extent during the Middle Ages, but the translations had not circulated widely, nor had they been very well understood nor appreciated fully. With the general improvement in mathematical literacy and humanist-sponsored reconstructing and re-learning of ancient mathematics, and the wider circulation—and then *printing*—of mathematical texts in the 15th and 16th centuries, there were more and more (though hardly many!) people who could understand Archimedes, and more who came to grips with his work and appreciated some of its implications. Most of the surviving (and then extant) works of Archimedes are severely pure mathematics: they represent one of the pinnacles of the Greeks' achievement in this area, and there were soon sixteenth century mathematicians who recognised this. But Archimedes *also* had a reputation for his mechanical inventions (burning mirrors, machines of war, water clocks, hydrostatic balance, planetarium, and so on), which makes him a little different from the standard pure, cerebral, "philosophical" classical mathematician. And further, there are two extant Archimedean treatises on the mathematics of static mechanical and fluid static phenomena: "On the Equilibrium of Planes" and "On Floating Bodies" respectively. The technical, foundational and intellectual power of these geometric proofs of the principles of statical phenomena was enormous. Archimedes fitted in with the kind of "theory of engineering" or science of the engineering arts that the humanist-inspired engineers had been doing for some decades—a "science" of engineering made scientific both by its use of mathematics and its origins in the works of the respected classical Greeks. Only Archimedes was very much more so. Reading Hero, Vitruvius, Pappus, and the rest gave these chaps an argument for the noble science of engineering (not that they would have put it this way); but Archimedes' reputation as a practical engineer *and* as a supreme pure mathematician made the power of the legitimisation of something like a mathematical science of engineering (or mechanics) vastly more convincing.

The technical competence that this ambitious, Archimedean-inspired mathematical engineering could give to the engineer is simple enough to understand: with the principles of fluid statics demonstrated in mathematical form, one could be much more certain how to load and rig and ballast ships, and with a better understanding of statics and the principles of the lever and the other simple machines, one could build larger and more powerful and complex machines (especially of war). Tartaglia, of whom more in a second, is a good example of this. But further, Archimedes did not just show the mathematics of the functioning of these machines, he demonstrated—proved geometrically—the mechanical principles that underlie these machines. This was critical in establishing the scientific status of mechanics. More of this in a moment.

Niccolò Tartaglia is one of the great unsung heroes of this story. He was not a particularly heroic personality—indeed, he is more easily classed amongst the "nerds" of history than the "heroes". But even nerds have their place in the cosmos. Tartaglia was not of high social status; quite the contrary. He started out life as a pretty much self-taught applied mathematician and teacher of various low level mathematical arts. However, he had visions of a better life for himself, and tried to take on the trappings of the humanist scholar and gentleman of learning. He presented his mathematics of the practical arts, and himself as a teacher and "thinker" about them, as developing the mathematical, mechanical sciences of the ancients, and giving rise to sure, certain, proven knowledge. In 1543 Tartaglia published a singularly archetypical work (and in a singularly auspicious year: 1543 also saw the publication of Copernicus' *De Revolutionibus* and Vesalius' treatise on anatomy *De Fabrica*), in which he published for the first time a translation of the two mathematical mechanical works of Archimedes, with a commentary. The translation, underneath the humanist rhetoric of his scholarly travails for his enlightened aristocratic patron, was stolen from the mediaeval translation, and is atrocious and full of errors: you should hear how some genuine humanist gentleman-scholars laughed at Tartaglia's Latin. But the commentary is an engineer's commentary, explaining how this theory can be applied, and enthusing on how this showed these applications of mathematics to the practical arts to give rise to sure and certain science, not to the mere conjecture and uncertain rules of the artificer.

Here and in other books, Tartaglia details how the Archimedean engineer is capable of heroic feats. One of the most cited examples of the time was the loss of a spanking-new ship of the Venetian navy just outside Venice. This new ship, just launched and fully rigged and armed, sailed out of Venice and promptly capsized. Not good for the health of the many sailors lost, and singularly embarrassing for the authorities to lose a ship just because it set sail. The *problem* was simple enough: too much high rigging and too much heavy cannon on the upper decks meant that the centre of gravity was just too high, and a little wind toppled the ship. I told you so, go the well educated engineers: if you follow Archimedes' teachings, you can calculate where the centre of gravity of a floating body is, and avoid these inconveniences. But the authorities faced a worse problem: this brand new ship lying in the outer harbour was in perfect condition—except that it was under water. The salvage teams were called in, but were unable to raise her as she had already become stuck in the mud. Then came the heroic engineers' solution: two heavily ballasted barges were brought alongside the hulk, many ropes attached to it by divers ... and then the ballast in the barges thrown overboard. Eventually, the force of buoyancy on the many ropes (vastly more force than one could get from a few cranes or winches) pulled the hulk out of the mud, enabling it to be salvaged. Easy if you have read Archimedes.

If Archimedes much improved the technical prowess of a mathematically literate engineer who read him, he also furnished powerful new technical arguments for the scientific status of this theory of a practical art. The use of mathematics in the sense of quantification did not change the status of a practical art: it remained an art. Consider a beam supported at either end. Place weights at the middle of the beam and eventually at a certain weight, the beam breaks: say at 100 kg. Clearly it *does* break at that weight, and not at 99 or 101 kg. But this is obviously a contingent truth: it could have (it didn't, but it *might* have) broken at some other weight. This is why engineering truths are not scientific truths. Similarly, consider a beam with weights M_1 and M_2 at either end, which balances about a pivot at some position. Clearly the beam balances there and not somewhere else—but it *could* have balanced somewhere else on the beam ... Which makes the behaviour of the balance contingent—and part of an art.

Except that if you have read Archimedes' proof of the law of the lever and the balance, a proof that comes out of geometry alone, then you now know that the beam *could not* have balanced anywhere else: you have a *proof* that it had to balance there and nowhere else. Science! Proofs from geometry (and what were to masquerade as geometric theorems) of the phenomena of mechanics make possible scientific knowledge of those phenomena—knowledge that is necessarily and demonstrably so.

Tartaglia was not the only Italian "ambitious humanist-Archimedean engineer" to start to draw these conclusions, but he was one of the earlier and more notorious. They were to draw the conclusion that since this mathematical mechanics treated the physical behaviour of bodies—of all inert beings—it must be some sort of underlying physical science: the first sign that it was the paradigm science. Furthermore, the next, post-Archimedean step to take was clear enough: if static phenomena could be expressed mathematically, could moving phenomena? Was it possible to establish a mathematical theory of both free fall motion (which was already part of the domain of science, as it is a natural occurrence) and the "violent" un-naturally caused motion of projectiles? Tartaglia was to have several attempts at producing a "new science" of projectile motion, extending the Archimedean analysis of statics to non-static phenomena ... and although it must be said his attempt is a complete failure, it at least shows that the possibility is on the ground (and in the late 1530s).

There were many other ambitious applied mathematicians over the middle and later sixteenth century, and they worked to extend the scope and the detail of this Archimedean mechanics. Perhaps the culmination of this story is a certain Guidobaldo del Monte, who published in 1577 a treatise on mechanics which was a more or less complete mathematical treatment of the five simple machines (lever/balance, pulley, axle/gear, wedge, screw) and their compounds, based on Archimedean geometric principles. After Guidobaldo's treatise, we can talk about an established—a demonstrated—mathematical science of mechanics; if you were interested in such a notion, it was on the table.

Guidobaldo, like his contemporaries, was a little ambiguous about the rôle of experiment in all of this: you need to do experiments to establish the behaviour and the principles of a mechanical phenomenon, but experimentation *per se* was not what made such knowledge scientific; that was a result of the mathematical proofs.

6. An Obscure Professor of Mathematics

It is convenient to end this story by looking at a friend and correspondent of Guidobaldo, a rather obscure lecturer in mathematics at Pisa (c.1589–92) and professor of mathematics at Padua (1592–1610). This chap, whom Guidobaldo liked, protected and helped—to get the jobs, first at Pisa, then at Padua—can be seen very much as a follower of Guidobaldo. Our professor worked on much the sort of things that Guidobaldo was doing, only he also started to try to extend this Archimedean mathematical mechanics to the phenomena of motion, of moving bodies and moving fluids. His own teacher of mathematics, Ricci, was himself a pupil of Tartaglia ... so the connexions are not difficult to see. Initially, it must be said, his studies of moving bodies and fluids were without great success, although at least his fidelity to what I would call this Archimedean mathematical-mechanical *tradition* was clear enough. He was, all in all, a typical example of this breed: hardly an artisan or of the artisan class, he held a Chair of Mathematics, frequented society (so far as he could) and worked hard to get himself a Chair in Philosophy (without success), as this was the next rung up in the academic ladder that he wished to climb, and scientifically much more prestigious than being a professor of mathematics ... better pay and working conditions, too.

In his twenty-odd year career as professor of mathematics, he, in common with other contemporary professors of mathematics at other Universities, would have taught on the mathematical theory of a wide variety of mechanical arts. These would include algebra, calculating astronomy, surveying, optics, military arts, civil engineering, design and use of mathematical instruments and so on: in short, the various empirical arts of engineering, treated from a mathematical point of view.

Our professor also earned a little bit of extra money on the side (mathematicians' salaries not being adequate, then as now) supervising undergraduates, and making various more or less useful instruments—with an eye to a commercial market. He made a light that works under water (very useful for the docks at Pisa), a hydrostatic balance and water-clock (following on from Archimedes), some optical devices, and the gadget that earned him most money, a very handy and useful artillerer's compass (a device for aiming and ranging cannon). Oh, and I almost forgot: he also heard about a neat optical device invented by a Dutchman that allowed distant things to be seen as if they were near by, and our professor realised that this could have lucrative military applications, made such a spy-glass, and sold it to the local military.

And then one day he turned this optical instrument to the heavens and made a series of stunning and unexpected discoveries: mountains and valleys and seas on the moon, implying the moon was not a perfect orb, as Aristotle taught, but an object much like the Earth; the phases of Venus, implying that Venus goes around the Sun and not around the Earth; and several moons going around Jupiter, thus demonstrating that the Earth was not the only centre of rotation in the universe—and thus destroying the classical physics of the heavens. All of a sudden, our rather obscure and—let us be frank—rather uninteresting Professor of Mathematics becomes an international figure, the foremost proponent of the Copernican model of the heavens. He suddenly becomes Galileo, and the Copernican model, heretofore somewhat ignored as simply a very thoroughly bizarre and physically untenable theory (do *you* feel the Earth moving?), gets some spectacular empirical support. Empirical support, however, that looks impressive to us, but to Galileo's contemporaries might have seemed less powerful: all these observations do not make the physics of the model any less impossible.

How do you make the model physically tenable? How do you account for the fact that the Earth does not seem to be moving? Galileo was able to construct a mathematical analysis of the mechanical phenomena of motion—both projectile and free fall—in his 1632 *Dialogue* that made it seem reasonable that the Earth could move and we would not know about it; that showed that the same mechanical laws govern both terrestrial and celestial motions. The details of the argument are less convincing (and "correct" in our eyes) than you might have been led to believe, but never mind: for all its faults, this mathematical-mechanically based argument seems to hold water. What is quite remarkable is that Galileo should have attempted to justify Copernicanism, or the physics of a Copernican model, on the grounds of mechanics. This says clearly that for Galileo, mechanics was a science; and he thought that he could convince his audience of the scientific nature of his argument using a mathematical mechanics.

The power of the Galilean argument in defence of Copernicanism was based on the acceptability of a mathematical and mechanical analysis of motion. His investigation of this mathematical mechanics was certainly based upon experimental work, and its certainty—and so its scientific status—was guaranteed by the ability to prove the mechanics with mathematics. Galileo was certainly able to put forward the most remarkable and powerful arguments in favour of a heliocentric model of the universe, but his methodology, his language of concepts and his physics ... all in all, the tradition of scientific thinking from which he came, stemmed from the tradition of intellectually and socially ambitious Archimedean engineering. A tradition which as a follower of Guidobaldo del Monte and as a Professor of Mathematics he was clearly firmly inside.

Which all goes to show that if you let the engineers learn some mathematics, they will break all the rules, rock the boat, and destroy sound, sensible and well-established science. And, you might care to note, engineers have carefully never been allowed to learn any mathematics since. One Scientific Revolution is enough.

G.H. Hardy and Control Engineering

Jonathan R. Partington

It is generally known that the Cambridge pure mathematician G. H. Hardy prided himself that his work had no applications to the real world. In his book *A Mathematician's Apology* (1940) he phrased it as follows:

> I have never done anything "useful". No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world.

However, these days, geneticists commonly use the Hardy-Weinberg theorem, a result predicting the distribution of offspring in a population; moreover even Number Theory has found application in the construction of ciphers. Hardy's discomfort would probably be completed by the knowledge that modern Control Engineers use the Hardy spaces of complex analytic functions, and that whole books have been written on a subject known as H_∞ Control Theory, where the H stands for Hardy.

The Hardy spaces date back to 1915, having been introduced in a paper of Hardy's in the Proceedings of the London Mathematical Society entitled *The mean value of the modulus of an analytic function.* They may easily be defined as follows: for $1 \le p < \infty$ define H_p to consist of those power series $f(z) = \sum_0^\infty a_n z^n$ which converge for $|z| < 1$ and such that the *norm* of f given by

$$||f||_p = \sup_{r<1} \left(\int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta/2\pi \right)^{1/p}$$

is finite.

The space H_∞ has a rather simpler definition: it consists of all power series bounded in the disc with norm

$$||f||_\infty = \sup_{|z|<1} |f(z)|.$$

Intuitively we think of H_p functions as being well-behaved inside the disc, but getting large near the boundary (but in a controlled fashion), so that they have boundary values which are p^{th}-power integrable, whereas the H_∞ functions can be extended to the circle and remain bounded. For example the function $f(z) = 1/\sqrt{z-1}$ is in H_p if and only if $p < 2$.

From the point of view of Systems Theory, it is more convenient to work with the Hardy spaces of the right half plane $\mathbb{C}_+ = \{x + iy : x > 0\}$, so that the Hardy spaces $H_p(\mathbb{C}_+)$ consist of functions analytic in the right half plane with p^{th}-power integrable boundary values, and $H_\infty(\mathbb{C}_+)$ comprises the bounded analytic functions.

Consider now a very simple linear system determined by the following differential equation:

$$\dot{y}(t) + ay(t) = u(t).$$

Here y is regarded as the output (unknown), and u is the input (given). One way of solving this equation is to take Laplace transforms, so that we define $Y(s) = \int_0^\infty e^{-st} y(t)\, dt$

wherever this integral exists, and similarly let U be the transform of u. The differential equation turns into an algebraic one, namely

$$(s+a)Y(s) = U(s)$$

so that if we can invert the Laplace transform we can solve for y. In general linear systems will always produce solutions of the form $Y(s) = G(s)U(s)$ for some function G.

Now it turns out in practice that under an assumption of finite energy, u will satisfy

$$\int_0^\infty |u(t)|^2 \, dt < \infty.$$

This is precisely the condition required to guarantee that $U(s)$ be in the Hardy space $H_2(\mathbf{C}_+)$. Moreover, to ensure that $y(t)$ is square integrable whenever $u(t)$ is, the condition is that $G(s)U(s)$ be in $H_2(\mathbf{C}_+)$ whenever $U(s)$ is in $H_2(\mathbf{C}_+)$, which is simply that $G(s)$ lie in $H_\infty(\mathbf{C}_+)$. These are the *stable* systems (so that we would require a to be positive in our original differential equation). However unstable systems are also a fact of life and these typically correspond to a function G with a finite number of poles in the right half plane.

Of course real life systems tend to have more than one input and output, and these correspond to spaces of matrix-valued functions. However, having observed certain Cambridge taxi-drivers, we are prepared to believe that one can sometimes do quite well merely by controlling one variable (such as the speed).

There is an easy measure of the complexity of a linear system using the notion of the *degree* of the system, whereby finite-degree ones correspond to rational functions $G(s) = p(s)/q(s)$ where p and q are polynomials, and the degree is the number of poles of the function, or equivalently the maximum degree of p and q. From an analytical point of view, infinite-degree systems are even more rewarding however.

It turns out that an appropriate measure of the 'closeness' of two systems lies in taking the H_∞ norm of their difference, so that the distance from $F(s)$ to $G(s)$ is given by $\|F - G\|_\infty$. Many problems of Control Theory can be reduced to optimisation problems using this norm—for example *sensitivity minimization* involves making sure that the output is as far as possible unaffected by noise in the input, and *model reduction* involves approximating a system by a simpler system (which is easier to build and analyse).

We conclude by offering a proof of the Riemann Hypothesis using Systems Theory, thus appeasing the ghost of Hardy by showing that Engineering may have applications to something useful (i.e. Number Theory), after all.

Let $\zeta(s)$ denote the Riemann zeta-function, defined by $\zeta(s) = 1 + 1/2^s + 1/3^s + \ldots$ for $\operatorname{Re} s > 1$ and extended uniquely to give a meromorphic function on the entire complex plane. We aim to show that the only zeroes of $\zeta(s)$ in the right half plane lie on the *critical line* $\operatorname{Re} s = 1/2$.

The reciprocal of the zeta-function is given by $1/\zeta(s) = \sum_{n=1}^\infty \mu(n)/n^s$, where $\mu(n)$ is the Möbius function of number theory, at least for $s > 1$, with a meromorphic continuation to the whole complex plane. It is therefore our concern to determine exactly where the poles of this function lie.

Now the mathematical reader may do this by using operator theory (there is a linear transformation—a Hankel operator—whose eigenvectors correspond exactly to the pole-positions); alternatively those readers with Meccano sets may care to construct a linear system corresponding to the function $G(s) = 1/\zeta(s)$. (We omit some details here through lack of space, using a well-known precedent due to Fermat.) The upshot is that given such a system we can work out its modes, and discover whether any of them correspond to poles that are not on the critical line. The reader will be reassured to know that in fact none of them do, so that the Riemann hypothesis follows.

Sir Michael Cometh

Sir Michael Atiyah is coming to Cambridge at the start of the new academic year. Sir Michael will take up residence as Master of Trinity, replacing the retiring Sir Andrew Huxley. The Archimedeans wish him well there, and also at the new Mathematical Institute, which he will head when it is founded.

This is not the first time Sir Michael's return to his *alma mater* has been mooted; it was in the air when the mastership last became vacant. The incident was recorded in *Eureka* **45** thus:

> Sir Michael Atiyah
> Was coming here,
> But the fellows of Trinity
> Showed no affinity.

The appointment was the subject of a charming but wholly fictitious story in the *Observer* newspaper, who linked it with the position held at the Universities' Funding Committee by Peter Swinnerton Dyer, another distinguished ex-Trinity mathematician and a contemporary of Sir Michael. Sir Peter is Chief Executive to the U.F.C.'s Chairman, Lord Chilvers, and is rumoured to be a continuing thorn in the side to the Prime Minister and to that extremely right-wing peer. According to the *Observer* story, Downing Street hinted that he might be appointed as Master of Trinity, and the fellows, detecting an ulterior motive behind the announcement, created. In fact when the appointment came due the Prime Minister had other things on her mind, having lost a couple of Cabinet ministers recently in a well-publicised manner, and was more than happy to rubber-stamp the first choice of the Trinity fellows. I am tempted to break into verse myself:

> Professor Sir Peter Swinnerton-Dyer
> Rose progressively higher and higher
> In the Universities Funding Committee,
> Which the P.M., in retrospect, thought was a pity.

Let the last word, however, go to John Greenlees, who penned the original clerihew about Atiyah. He writes:

> Sir Michael Atiyah
> Was opposed by a peer,
> But the fellows of Trinity
> Weren't awed by nobility.

Unfair Games

Martin Baxter

"Fair is foul, and foul is fair", cry the witches in *Macbeth* warning against reliance on surface appearances, a recurring theme throughout the play. Taking the theme, here presented are some games, which themseleves are not as fair as they might seem. The skill of each game lies not in its playing but with the analysis and understanding of it, for we shall see that the outcome (even of random games) was never in doubt—but foully predetermined.

BRUSSELS SPROUTS: A topological unfair game

Many readers will be familiar with the two player game of *Sprouts*, in which every move extends a graph. (A graph is a set of points (vertices) joined together by non-intersecting lines (edges) on some topological surface, such as a sheet of paper.) Starting with two unconnected vertices, a move consists of drawing a new edge from one existing vertex to another, or even back to itself. The only restriction is that no vertex can have more than three edges connected to it (counting double for looping back edges). The new edge cannot cross itself or an existing edge. The move is completed by creating a new vertex half-way along the new edge, splitting it into two edges, and leaving the new vertex with the capacity for only one more edge to be connected to it. Players move alternately, and the winner is the last player able to make a valid move. Figures 1 and 2 are two sample games:

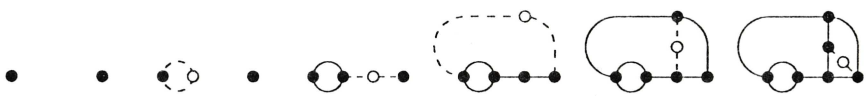


FIGURE 1. Player 1 wins.

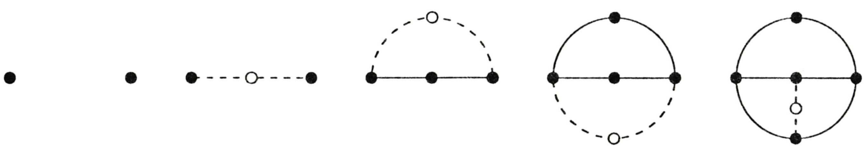


FIGURE 2. Player 2 wins.

By noting that the total capacity of the graph to accept new edges decreases by one with each move, we see that the game must finish in five moves or less. The game can be made longer and more complex by increasing the initial number of vertices. In fact the 2-sprout game is (fairly) easily analysed and an intelligent player 2 will always win.

In *Brussels Sprouts* we replace the dot shaped vertices with cross-shaped ones (+), and insist that all edges be attached to a free point of some cross. The new vertex placed on a new edge should have one bar of the cross lying along the line, and a cross-point lying on either side. Figure 3 demonstrates a game with one initial vertex:

FIGURE 3. Player 1 wins.

It is not immediately apparent that this does not just complicate the game further (it certainly lasts longer), but in fact the analysis is greatly simplified. This is because the game is determined by the faces of the graph. A face is a connected region of the space closed off from the other faces by the edges of the graph. We also consider the region of space outside the graph as a face. Each face must have at least one free cross-point in it, as the newest boundary edge of the face left a free cross-point on each side. If there is a face with more than one free internal cross-point, then the game is not yet over as one valid move is just to join them and split the face into two. So the game ends when there is exactly one free cross-point in each face. Each move uses up two cross-points, but also creates two new ones, leaving the total unchanged. Thus in the 2-sprout game, the game ends whenever there are 8 faces of the graph. To discover when this is, we use *Euler's Formula*:

$$V - E + F = 2,$$

where V = number of vertices, E = number of edges, F = number of faces. (The precise conditions for validity will be discussed later.) After n moves these values will be:

$$\begin{aligned} &V_n = n + 2 && \text{(each move adds a new vertex)} \\ &E_n = 2n && \text{(each move creates a new edge which splits into two)} \qquad (1) \\ \text{Thus}\quad &F_n = n && \text{(by Euler's formula)} \end{aligned}$$

We know that the game ends when and only when $F_n = 8$. Thus (by (1)) the game always ends after exactly 8 moves, and player 2 will always win, no matter what either player does during the game. Try this for yourself now for a couple of games; you will always finish with 10 vertices and 8 faces.

Now if we start with m vertices, then:

$$V_n = n + m, \qquad E_n = 2n, \qquad F_n = 2 - m + n.$$

The game ends when $F_n = 4m$, that is when $n = 5m - 2$. So player 1 wins when m is odd, and player 2 wins when m is even. It is easy to confuse the unwary opponent by changing the parity of m if he or she becomes suspicious of the order of play. The situation can be further complicated by having k players instead of two, the winner being the last to make a valid move. Then player n will win if and only if

$$n \equiv 5m - 2 \pmod{k}$$

Luckily 5 is prime, so for any k not divisible by 5, there is a range of values of m to make any given player the winner.

The *coup de grâce* is dealt by playing the game on more interesting toplogical spaces than flat paper, such as the torus, Möbius strip, sphere, and so on. These surfaces are characterised topologically by their number of "handles" (*genus*), and their number of "crosscaps". The generalised Euler formula is:

$$V - E + F = 2 - 2g - h,$$

where g is the number of handles, and h the number of crosscaps (Möbius strips sewn into the surface). The following remarks on the validity of this formula may safely be ignored.

* * * * *

Strictly Euler only applies to a graph embedded in a closed combinatorial surface. So the graph must be connected, and all its faces simply connected.

(1) The final graph will be connected, as any two components would have a free cross-point in the outer region, which could be joined as a next move.

(2) The knowing player must ensure by the end of the game that all the faces are simply connected, as there is no guarantee that this need happen otherwise. In practice this means using all the wrapovers and points of commonality discussed later. For example, on the torus there should be edges going round the circles shown in Figure 4.

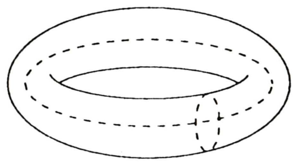


FIGURE 4. Torus.

(3) It is certainly true for connected closed combinatorial surfaces, which are exactly the connected compact 2-manifolds. And these, apart from the sphere, are either a sphere with g handles, or a sphere with h crosscaps. (A sphere with g handles and h crosscaps is equivalent to one with $2g + h$ crosscaps, $g, h \neq 0$.)

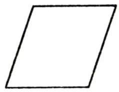
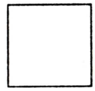
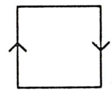
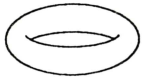
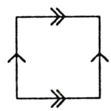
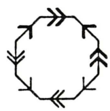
Euler is also valid for C (the complex plane), its open connected subsets, cylinders, and many other two dimensional surfaces, as they can all be mapped onto the 2-sphere for which Euler is valid. Table 1 describes some "common" surfaces.

* * * * *

So at the end of the game:

$$n \equiv 5m + h + 2g - 2 \pmod{k}$$

With three degrees of freedom (m, g, h), a wise huckster could keep opposition confusion going for an impressive amount of time playing with rotten sprouts.

Table 1. Surface summary.				
Surface	Picture	Representation		g, h
C-plane, sphere, etc.			square	0,0
Möbius strip			square with one pair of opposite sides identified	0,1
Torus			square with opposite sides identified	1,0
Double torus			octagon formed by joining two tori	2,0

NIM: unfair in binary

Nim has been understood since Charles Bouton's paper of 1901, so there is nothing new in this account, which is included for completeness and as an illustration.

In Nim there are a number of rows, each with a number of counters along it. A traditional starting pattern is to have n counters in the n^{th} row (Figure 5), but this does not matter much.

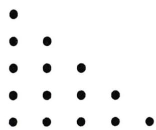


Figure 5. *Nim* start position.

A move is simply to choose a row, and to remove from it as many counters as desired. The winner is the player who removes the last counter. There is a suprisingly simple function of the position which reveals which player can win.

Many will know the logical binary operators AND ($\wedge$) and OR ($\vee$), perhaps fewer the operation EOR ($\veebar$) (or XOR)—the exclusive-OR. Its truth table is shown in table 2.

Table 2. Exclusive-OR		
a	b	$a \veebar b$
0	0	0
0	1	1
1	0	1
1	1	0

We can extend EOR (and the other operations) from acting only on $\{0,1\}$ to acting on all non-negative integers, via their binary expansions. That is:

$$\text{If } a = \sum_{n=0}^{\infty} a_n 2^n \quad b = \sum_{n=0}^{\infty} b_n 2^n, \quad \text{then } a \veebar b = \sum_{n=0}^{\infty} (a_n \veebar b_n) 2^n$$

(In fact $(\{0,1\}, \veebar, \wedge)$ is just the field F_2, and we create the commutative product ring $(\{0,1\}^{\mathsf{N}}, \veebar, \wedge)$ under componentwise operations, the non-negative integers being a subring. So $(\{0,1\}^{\mathsf{N}}, \veebar)$ is an abelian group of elements of order 2.)

Returning to Nim, if we represent the state of the game as $r_1, r_2, \ldots$ where r_n is the number of counters left in the n^{th} row, then we consider $R = r_1 \veebar r_2 \veebar \ldots$ What does it mean when $R = 0$?

(i) There are at least two non-empty rows, so the player about to play is not just about to win.

(ii) Once that player has moved, R will no longer be zero. For if he reduces row n from r_n to r'_n counters, then R changes to R':

$$R' = R \veebar (r_n \veebar r'_n) = r_n \veebar r'_n \neq 0$$

(iii) We shall prove shortly that whatever the move was, there is a move by the other player which returns R to zero.

So if player 1, say, can arrange for R to be 0 after his move, then player 2 cannot win with his next move and player 1 can return R to 0. Thus player 2 can never win, so player 1 must win (as the game always ends).

Let us quickly see how to return R to 0:

Let k be the position of the leading "1" in the binary expansion of R (non-zero). Then there exists an n such that r_n has its k^{th} bit set, as R indicates that there is an odd number of such n. Set r'_n to be $R \veebar r_n$, then the bits of r'_n higher than the k^{th} are unchanged as R is zero there, and the k^{th} bit itself is zero. Thus $r'_n < r_n$, so the move to make is to remove $r_n - r'_n$ counters from the n^{th} row, as $R' = R \veebar (r_n \veebar r'_n) = r'_n \veebar r'_n = 0$.

The above shows how to make R equal to zero given any position in which it is not. So a knowledgeable player 1, if the initial value of R is non-zero, will win whatever player 2 does. Similarly a knowledgeable player 2 will win if R is initially zero. If one of the players is unknowledgeable he or she will probably leave R in a non-zero state after some move, enabling the other to zero R and take the advantage.

We see that the game is determined by the initial value of R, R_0. For a starting position of n rows with i counters in the i^{th} row, $R_0 = \veebar_{i=1}^{n} i$. Now $(2m) \veebar (2m+1) = 1$ so we deduce:

$$R_0 = \begin{cases} n & \text{if } n \equiv 0 \bmod 4 \\ 1 & \text{if } n \equiv 1 \bmod 4 \\ n+1 & \text{if } n \equiv 2 \bmod 4 \\ 0 & \text{if } n \equiv 3 \bmod 4 \end{cases}$$

Player 2 can win when 4 divides $(n-3)$, player 1 otherwise.

In a sense neither Nim nor Brussels Sprouts is any more unfair than any other discrete (turn-about) two-player game, such as chess. For, as long as the game always ends, there will be a strategy for one of the players which gives victory no matter what the other does. In Brussels Sprouts we saw that the winner is determined by the initial number of vertices and the winning strategy is arbitrary, as it does not matter at all. In Nim the position is characterised by the value of R, and the simple winning strategy is described above. In chess, say, there are 3 outcomes (win, lose, or draw) yet an optimal strategy for either black or white does exist. (History suggests that it could be white who is always able to force at least a draw, but it is an open question.) However it is doubtful whether it is governed by some simple rule or formula and the computation involved in calculating it (and it is computable in a technical sense) would be substantial—there are of the order of 10^{50} positions.

DOUBLE OR QUITS: walks around random examples

Now we consider the result of a random game to be the (real-valued) net gain X of total rewards or winnings less any stake, penalty or cost due to playing the game. If the game is voluntary, with gain Y due to not playing the game, we shall consider $X - Y$, the relative gain of playing the game. Knowing the distribution of the net gain X, we want to know if the game is worth playing, or at least fair. Two simple alternative questions to ask are: is there an evens chance or better of at least breaking even? and do we break even on average? In probabilistic language:

$$\text{Is } \mathsf{P}(X \geq 0) \geq \tfrac{1}{2}\,? \qquad \text{Is } \mathsf{E}(X) \geq 0\,?$$

(We are assuming that X is an integrable random variable on some probability space.) These distinct conditions are often confused. For example, even Adam Smith asserts of a lottery that:

> There is not, however, a more certain proposition in mathematics than that the more tickets you adventure upon, the more likely you are to be a loser.
>
> Adam Smith, *Wealth of Nations*, I.x.i

Consider a lottery of 1000 tickets each at unit cost, with only one prize of 900. Buying n tickets, with a resultant net gain of X_n, we see that:

$$\mathsf{P}(X_n \geq 0) = \begin{cases} \frac{n}{1000} & \text{if } \ 1 \leq n \leq 900 \\ 0 & \text{if } 900 < n \leq 1000 \end{cases} \qquad \text{but} \qquad \mathsf{E}(X_n) = -\frac{n}{10}$$

So the probability of being a net winner increases with the number of tickets bought (up to a limit), but the expected gain decreases. As no-one ever made a living by winning lotteries, one might assume that the expectation is the indicator that the wise should use. One could further ask why so many people bet on horse races, play slot machines, and enter the football pools; all activities with negative expectations of gain. Prof. Blimp would inform you that these people have no head for such things and have conquered rationality by a combination of greed and the natural human belief in one's own good fortune. His sociology colleague Dr Larebil suggests that what is being sold is not just the chance of a fortune but the opportunity to dream and to hope about winning it. One former Chancellor of the Exchequer, known to enjoy playing the football pools, clearly agrees:

> **Lawson:** Well, the sum you might win is absolutely enormous.
> **Question:** But the odds are enormous too.
> **Lawson:** Yes, I know, but it's worth doing just for the money, just for the possibility of getting this enormous sum.
>
> William Keegan, *Mr Lawson's Gamble*, Hodder and Stoughton 1989

This dream price can be the whole cost, as in Big Brother's lottery for the proles in Orwell's *1984*, which never really awarded the large prizes that it claimed to. Big Sister's premium bonds at least satisfy $\mathsf{P}(X \geq 0) = 1$ and $\mathsf{E}(X) > 0$ for a unit stake over one year. In fact $\mathsf{E}(X) = 0.07$ (7% interest rate), which implies a negative relative gain, when compared with investing in a bank or building society with $\mathsf{E}(X)$ nearer 0.09 (or 9% net interest).

However Prof. Blimp takes care each year to renew the insurance of his house against fire, a game where he is certain to lose his stake (premium) each year for no net gain, that is $X = -1$ almost surely. But, if he did not play the game, that is if he failed to insure his house, his gain Y, would be 0 with a high probability, but would be minus the value of his house with the small (but positive) probability that he leaves the gas on, or some other conflagrationary accident occurs. So his relative gain $X - Y$ is probably slightly negative, but has a chance of being very large and positive. As the insurance company has to make provision for overheads and profits, the premium will be more than the value of the house multiplied by the probability of its incineration: $\mathsf{E}(X - Y) < 0$. We now see that the relative gain of playing the football pools has exactly the same stochastic structure as the relative gain of insuring a house.

I am not however advocating either the purchase of 1000 pools coupons each week, nor the cancellation of insurance policies, and neither will *Eureka* accept any liability for any losses incurred as a result of such action—for there is a difference between the two cases. In the former, you pay to have a tall but probability-thin spike added to your "gain-space", in the latter you pay to have a similar, but negative, spike removed. In one the maximum gain is made much greater, in the other the maximum loss is made much less (the minimum gain is made much greater). The examples are here to show that, in real situations, even the stochastic structure of the gain random variable may not be enough to decide the question.

To conclude, I present a general case where the probability of success and the expectation of reward differ in a staggeringly paradoxical fashion.

Consider a sequence of random win/lose games against a 'house' or bank. The player only chooses the stake which linearly scales the gains and losses of the game. The stake can be zero (but not negative), allowing the player to stop playing at any time, by setting the stake to be zero from then on. The stake can depend on the results of previous games, but cannot depend on the results of games yet to come—the bank does not allow betting after the event. The games are probabilistically independent, with the n^{th} game yielding probability p_n of success with a reward of $\alpha_n > 1$ for a unit stake. So the gain X_n, of playing game n is:

$$X_n = \begin{cases} \alpha_n - 1 & \text{with probability } p_n \ (0 < p_n < 1) \\ -1 & \text{with probability } q_n = 1 - p_n \end{cases}$$

$$\text{and} \quad M_n = \sum_{r=1}^{n} C_r X_r$$

is the total gain so far, where C_n (≥ 0) is the stake placed on game n. The *Borel–Cantelli Lemmas* imply that the probability of the player having infinitely many victories is either 0 or 1, depending respectively on whether the sum $\sum p_n$ converges or diverges. Unbelievable paradoxes arise in either case.

Case 1. $\sum p_n < \infty$, *certain to have only finitely many victories.*

Suppose $\alpha_n = \frac{2}{p_n}$, then $\mathsf{E}(X_n) = 1$, and we choose to gamble a unit stake each turn, $C_n = 1$, then we see:

$$M_n \to -\infty \text{ with probability 1} \qquad \text{but} \quad \mathsf{E}(M_n) \to +\infty$$

Case 2. $\sum p_n = \infty$, *certain to have infinitely many victories.*

The gambler can choose each turn to stake exactly the amount needed to restore all past losses and give a small profit if victorious in the game:

$$C_n = \frac{n - M_{n-1}}{\alpha_n - 1} \qquad \text{so } M_n = n \text{ if successful.}$$

This results in a fortune strictly increasing from one victory to the next, and if the gambler chooses to stop after the k^{th} victory (and there will be a k^{th} victory with probability 1), then he finishes with a positive net gain. A slightly simpler case is $C_n = (N - M_{n-1})/(\alpha_n - 1)$ which stops after the first victory with a profit of N, for arbitrary 'greed value' N. This is a generalization of the stake doubling rule commonly associated with roulette, under which the initial bet is 1, which is doubled after every losing spin of the wheel, so that the eventual victory yields a net gain of 1. In this case $p_n = q_n = \frac{1}{2}$, $\alpha_n = 2$, and $C_n = 2^{n-1}$ until victory. By picking

$$\alpha_n = \frac{p_n + 1}{2p_n} \qquad \text{we get} \quad \mathsf{E}(X_n) = -\tfrac{1}{2}q_n, \text{ so } \mathsf{E}(M_n) < 0.$$

So we have a negative expectation of M_n, yet M_n itself tends to some positive value N with probability one.

In both these cases, the limiting value of M_n gives the true worth of playing the game, and the expectations are misleading. The games are not however as good for the theoretical winner as might be supposed. In the former, where the bank wins, the amount "invested" in paying out to the player on his finite number of wins has infinite expectation, so the bank needs deep pockets. Alternatively if only a score is being kept, the real winner is the player who does not keep score, as an infinite amount of paper will have to be bought to record it.

In the latter game, in the simple stop-after-first-win case, the expected losses before the first win can have finite expectation. (If $\alpha_n = 1/p_n$, then $\mathsf{E}(M_n) = 0$ and $\mathsf{E}(|M_n|) \leq N$.) The only snag now is that although $\mathsf{P}(T < \infty) = 1$, where T is the time of the first victory, it may be that $\mathsf{E}(T) = \infty$, that is you should expect an infinitely long wait before success. For example,

$$p_n = \frac{1}{n+1}, \quad \mathsf{P}(T \geq n) = \frac{1}{n}, \quad \mathsf{E}(T) = \sum_{n \geq 1} \mathsf{P}(T \geq n) = \infty.$$

REFERENCES AND ACKNOWLEDGEMENTS

Sprouts: I am indebted to Eric Brown of Moray House College of Education in Edinburgh who introduced me to topology via Brussels Sprouts, and to Andrew Wren of Cambridge for checking my topological statements.

Nim: More on Nim can be found in Chapter 15 of

Martin Gardner, *Mathematical Puzzles and Diversions*, Pelican, 1965.

Random games: is composed of snippets stolen from lectures, example sheets, and other sources.

A Logical Problem

Colin Bell

[This was the first problem which candidates entering the 1989 Puzzle Hunt were required to solve. For a few further details of the Puzzle Hunt see page 20.]

The Archimedeans are bankrupt. As a creditor, still unpaid for fifty packets of Jaffa Cakes supplied, you must find out who is the Junior Treasurer. You have assembled five committee members before you, denoted A, B, C, D and E, whom you know to be the President, the Vice-President, the Secretary, the Junior Treasurer and the Registrar in some order. They have a scheme to stop you finding out which of them has which post. Of course, they know which posts the others hold.

If the last statement made about a given person is true, then he will lie. If the last statement made about him is false, he will tell the truth. If no statement has been made about him, then he can do either.

(NB: The statement "A is the President" will affect A's subsequent statements, provided no statement about A is made in the interim, but not those of the President.)

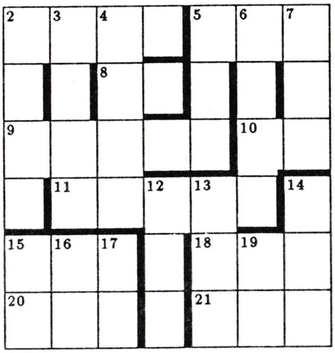
They make the following statements (in this order):

A: D is the Vice-President.
B: C is the Secretary.
C: A is the Registrar.
D: B is the President.
E: C is the President.
A: B is the Junior Treasurer.
B: C is the President.
C: E is the Vice-President.
D: A is the Junior Treasurer.
E: D is the Vice-President.

The solution may be found on page 92.

Debased

Quinapalus



An answer with index n is to be entered in the grid in base n. No answers have leading zeros. Use the letters A to K to represent the digits ten to twenty. The vertical symbol | means "divides exactly into".

Across	Down
2A | 5A | 9A	2D | 3D | 4D | 5D | 6D | 7D
8A | 10A | 11A	2D | 17D
15A | 18A	5D | 15D
15A | 20A	6D | 12D | 13D | 14D
	16D | 19D

On Large Numbers

Mark Wainwright

The following is a rejuvenated version of an old paradox. In its original form it goes as follows:

GAME 0. I give you two indistinguishable envelopes, and tell you that each contains a cheque for a positive sum of money, one of these being twice the other—y and $2y$, say. You may open one envelope and examine its contents, and then choose which to take. Your aim is to maximise your profit.†

Now, you open an envelope—label it A, say‡—and find it to contain an amount which we will also confusingly call A. You reason as follows: envelope B contains $\frac{1}{2}A$ with probability $\frac{1}{2}$, and $2A$ with probability $\frac{1}{2}$. So the expected value of G, your gain from swapping, is

$$\mathsf{E}(G) = \tfrac{1}{2}(2A + \tfrac{1}{2}A) - A = \tfrac{1}{4}A \tag{1}$$

which is positive. Thus you should always swap.

On the face of it this looks ludicrous, as indeed it is. The usual counter-argument runs as follows: in assuming that the cases $B = \frac{1}{2}A$ and $B = 2A$ are always equally likely, we have tacitly assumed that the amount y has a probability distribution which is uniform on $\mathbf{R}$. But no such uniform distribution exists (it has to integrate to 1), so there's no problem.

It would be nice to show that there is *no* probability distribution which enables us to generate the paradox. To this end, we note that (if the paradox is generated)

$$\begin{aligned} \mathsf{E}(B|A = a) &> a \qquad \forall a \\ \mathsf{E}(B - A|A = a) &> 0 \qquad \forall a \end{aligned}$$

By the Tower property of conditional expectation§

$$\sum_a \mathsf{E}(B - A|A = a)\mathsf{P}(A = a) = \mathsf{E}(B - A),$$

where the sum is over all possible values of A. In the continuous case the sum becomes an integral, and in any event we get the same result, *viz.* that $\mathsf{E}(B - A)$ is strictly positive, so $\mathsf{E}(B) > \mathsf{E}(A)$. But we can run the argument through again transposing A and B, and find that $\mathsf{E}(A) > \mathsf{E}(B)$, a contradiction. So we can't generate the paradox after all.

---

† It is irrelevant to this article, but it happens that even here you actually have a sensible strategy. Indeed more generally, if the envelopes contain distinct but otherwise arbitrary real numbers and your aim is to end up with the greater, you have a strategy which succeeds strictly more than half the time. I leave it as an infuriatingly difficult exercise for the interested reader to find what this method is.

‡ For convenience I shall always suppose envelope A to be the one you open first, and will label the other one B.

§ which is intuitive and easily verified from the definition of conditional expectation.

This was all very well until Marcus Moore proposed the following:

GAME 1. This is much the same as Game 0. However, I give you the added information that the amounts in the envelope are powers of 2, say 2^N and 2^{N+1}, where N has been chosen according to

$$\mathsf{P}(N = n) = \epsilon(1-\epsilon)^n, \qquad n \in \mathsf{N} \tag{2}$$

where N is the natural numbers including 0, and I reserve the right to be more specific about ϵ later, but I shall say now that it is in $(0,1)$ so the probabilities sum to 1 as they must do. Now if you find that $A = 1$, you know $B = 2$, so you swap. Otherwise; suppose $A = 2^n$. Then plugging in the formulae we find

$$\mathsf{E}(G) = \frac{2^{n-1}}{2-\epsilon}(1-2\epsilon), \tag{3}$$

which is positive, as long as $\epsilon < \frac{1}{2}$.

Readers who have read Martin Baxter's article elsewhere in this issue might say with a knowing wink, "Ah, but expectation is a load of misleading bunkum", or words to that effect. I urge such readers to consider the case, say, $\epsilon = \frac{1}{3}$, and to consider cases. For example, in this case, we get the pair $\{1,2\}$ with probability $\frac{1}{3}$, and $\{2,4\}$ with probability $\frac{2}{9}$. So in every 9 turns we would expect to have roughly the following pairs:

$$\begin{array}{ccccccccc} 1 & 1 & 1 & 2 & 2 & \mathbf{x} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ 2 & 2 & 2 & 4 & 4 & \mathbf{x} & \mathbf{x} & \mathbf{x} & \mathbf{x} \end{array}$$

where none of the $\mathbf{x}$'s is 2. So if you play a large enough number of times, you mutter about the Strong Law of Large Numbers and expect to be gaining an extra £1 for every five times you find a 2 as long as you swap—because $1+1+1+4+4 = 11$ whereas 5×2 is only 10. Notice that this gives $\mathsf{E}(G|A = 2) = \frac{1}{5}$, which agrees with (3). The reader may easily convince himself that this is no coincidence, and that the same argument will work for any value of A.

Now we showed above that we couldn't have a situation where $\mathsf{E}(G)$ was always positive, and now we've shown that indeed we can. Luckily there is no real contradiction. We needed then to show that $\mathsf{E}(A) > \mathsf{E}(B)$. But the cases we're considering (i.e. $\epsilon < \frac{1}{2}$) turn out, fortunately, to be precisely those in which both these expectations are infinite; so the argument falls down. At first it looks as though $\mathsf{E}(B - A)$ should still be defined and positive (in fact, the sum we would be most likely to write down turns out to be infinite), but luckily a little analysis comes to the rescue here. The sum which appears to consist entirely of positive terms is in fact the sum of *two* infinite series, giving the positive and the negative terms for each value of n (i.e. the gains and the losses). Neither series converges absolutely; so the fact that we have found an arrangement giving an infinite sum is quite meaningless: this sum doesn't tell us anything, as we could quite easily rearrange it to give any answer we like.

On the other hand, it is still unsatisfactory that, having opened the first envelope, you always expect the other envelope to contain more. In order to make the problem more acute, let's play:

GAME 2. I've written two cheques as before. You open the first envelope, and then have the option of continuing, which consists of paying me the amount written in the envelope you've seen and taking the cheque in the second envelope

This elegant refinement is due to Richard Tucker. I'm not interested for the moment in the fact that there is obviously a genuinely sensible strategy for you, e.g. play only when $A = 1$. What matters is that, in every case, your *optimal* strategy appears to be to play. Richard was at first all in favour of making you decide whether to play before you've seen the contents of the first envelope; but that's a hopeless line, because then you're palpably being taken for a ride. By any reckoning, before you open either envelope, you must expect your gain to be zero if you expect it to be anything at all.

Now you might think "Yes, my expected gain is positive, but the total expected amount of money changing hands is infinite; so we're just subtracting infinity from infinity". But that won't do, because *after* you open the first envelope everything is irritatingly finite, and the arguments go through as before. In fact notice (from (1)) that $\mathsf{E}(G)$ is increasing for $n \geq 1$, so in particular it's clearly uniformly positive; in fact, again taking for definiteness the case in which $\epsilon = \frac{1}{3}$, you always stand (i.e. expect) to gain at least $\frac{1}{5}$, which is to say 20p. With this in mind we'll play another game, for which I am again indebted to Richard Tucker:

GAME 3. The scenario is as in Game 2, with $\epsilon = \frac{1}{3}$; but if you decide to play, you pay me 10p on top of the amount in envelope A.

Now, predictably, at each play you go through the probabilistic stuff, mumbling about expectation or the Law of Large Numbers as before, and find that you stand to win at least 10p—so you play. While for my part, I appear to have found a 10p generator, since manifestly your policy of always taking the second envelope can't *really* be any better than always taking the first envelope. Given that the expected amount of money in each envelope is infinite this may sound a little dodgy (actually I hope it doesn't; you can't really expect to gain by your strategy?) but luckily, for my sanity at any rate, it isn't. Specifically, remember I wrote the cheques so I know the amounts in the envelopes, and so clearly I expect to be winning 10p a shot. In fact, whatever else happens, I'm not going to win exactly 10p on any one go; that's just what happens on the average. To make it excruciatingly rigorous (here I mutter about Large Numbers myself), if we play sufficiently often, for every two times I put £2^n and £2^{n+1} in the envelopes, I expect you to pick each amount once, netting yourself a loss of 20p.

At this point the reader may be inclined to do a number of ill-advised things. He might, for example, go carefully over all the bits of maths. he skated over the first time through, and find, with increasing despondency and alarm, that they all work; and, this done, leaving a brief suicide note and leaping from the window might seem an attractive option. I urge such readers to exercise restraint in the latter respect, at least, until they have read on sufficiently far to discover the catch.†

For there is, of course, a catch. Equation (2) tells us that there are possible amounts in the envelopes which are arbitrarily unlikely. So however many times you play, there will be some (pairs of) amounts which will come up sufficiently infrequently that the Law of Large Numbers does not apply. The most ill-behaved numbers in this respect are the largest amounts, and hence much the most significant, especially as the Law of Large Numbers will see to it that most of the smaller amounts (probably) cancel out. We know that if you play more and more, the maximum amounts for which I have to

† Of course it would generally be impossible to leap from the window without *first* discovering the catch, at any rate if the window was initially shut. In order to obviate any possible confusion, therefore, I should make it clear that it was not to this sort of catch that I am here referring.

write cheques will also grow (with probability 1). And who is winning at the end of the day depends precisely on which envelopes you open on those few occasions.

So in fact there is no limiting behaviour on this game as we play again and again, so we can't make sensible remarks about what happens in a single game. This may seem unsatisfactory. "That", I hear you remark, "is all very well for aggregates of games, but what about the calculations which showed that, *after I'd opened the first envelope*, I'm always better to swap? They were all based on entirely finite, reasonable calculations, so even if they don't work in practice, how do they square with the fact that I'm obviously not doing any better that way than by keeping my first choice?"

The answer is that the calculations really were correct. If you play enough times, you *will* be ahead overall on the occasions when $A = 2$ if you do swap, rather than if you don't. Similarly for when $A = 4$, and so on. That is to say that

$(\forall \epsilon \in (0,1))\,(\forall n \in \mathsf{N})\,(\exists m \in \mathsf{N})$ after m games you're winning on the occasions when $A = 2^n$ with probability $\geq 1 - \epsilon$.

What we want to do is to change the order of the last two qualifiers; which we aren't allowed to do, otherwise Analysis II would be an even more pointless course than it already is. However many times you play there will always be a few large amounts which don't balance out properly, and on the basis of those few but unpredictable amounts are you winning or losing at any moment.

What you should gather from this—what I would like to convice you of—is that the Strong Law of Large Numbers really is a strong and important result. We may think that if something has probability $\frac{1}{2}$, that means it'll happen about half the time; that, we could add, is what probability means. But it's nothing of the kind, and in fact it isn't even true if we don't have the Strong Law to tell us it is. Trying to get anything sensible out of this game is precisely trying to do probability without the Strong Law, and any results that may come out are just so much waste paper. You can't do it. It doesn't make sense.

This gives rise to a final and rather lovely series of observations: if you decide on a cut-off point more than which you aren't prepared to lose in a single turn (we'll talk about Game 2 to make life easy for ourselves), then everything converges properly and you are winning on average. This is obvious (as I remarked earlier) if you aren't prepared to lose anything at all. Then you swap only when you find $A = 1$, and the expected gain from this strategy is precisely $\frac{1}{2}\epsilon$, since $A = 1$ with this probability. But you could choose (say) 2^{1990} as the maximum value of A for which you'd swap, and a similar analysis applies. This is at first surprising, because one might think that such a strategy is for all practical purposes equivalent to the bad old policy of always swapping; as it (probably) is for short series of games. In fact if you decide to swap only when $A \leq 2^n$, then your expected gain obviously vanishes except when $N = n$, and your overall expected gain per turn is

$$\tfrac{1}{2}\epsilon(1-\epsilon)^n 2^n$$

$\epsilon < \frac{1}{2}$, so $k = 2(1-\epsilon) > 1$ and $\mathsf{E}(G) = \frac{1}{2}\epsilon k^n$ which is monotonically increasing, a rather charming result. It is lent its piquancy by the unpleasing stochastic nature of the game when large amounts are changing hands. As a result of this, the larger your cut-off point, the longer you must play to be sure that your expected winnings will manifest themselves.

The Wobbler

Frederick Flowerday and David Singmaster

1. Abstract / introduction

The Wobbler is a geometric device invented by the first author some years ago. It consists of two interlocked circles. The planes of the circles are orthogonal and the centres of the circles are separated by $\sqrt{2}$ times the common radius. In practice, it is made from two circular discs, each with a radial notch. When the Wobbler is rolled on a plane, it executes a most fascinating wobbly motion. Observation indicates that the centre of gravity remains at, or nearly at, constant height. We demonstrate that it does remain at constant height and that this holds if and only if the separation of centres is $\sqrt{2}$ times the radius. We also study the distance between the points of contact and show that this is constant if and only if the separation of centres is equal to the radius. Such a device was invented by Paul Schatz [2].

There are many other natural questions which arise concerting the Wobbler and on which we have not been able to make any progress. We pose these for further investigations.

2. The height of the centre of gravity

Let the discs have radius r and let the separation of centres be $2d$ (Figure 1).

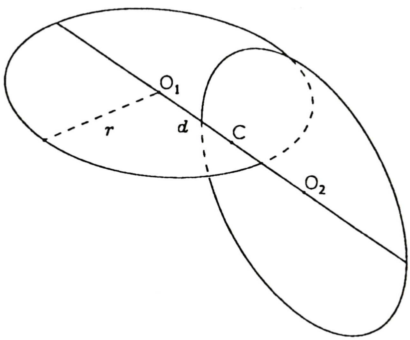


FIGURE 1

Clearly the centre of gravity, C, will be D from each of the centres which we denote O_1 and O_2.

There are two natural positions of the Wobbler In Position 1, each disc is at 45° to the base plane. An end view is shown in Figure 2. Then the height h_1 is $r/\sqrt{2}$ and the relative height η_1 is $1/\sqrt{2}$. In Position 2, one disc, say the second, is perpendicular to the base plane, as in Figure 3. Let A_1, A_2 be the points of contact. Then $A_1C = r + d_1$, $A_1O_2 = r + 2d$, $O_2A_2 = r$, so that $\eta_2 = h_2/r = (r + d)/(r + 2d)$.

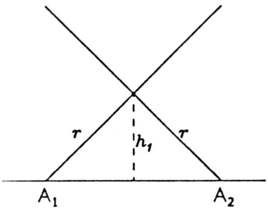


FIGURE 2

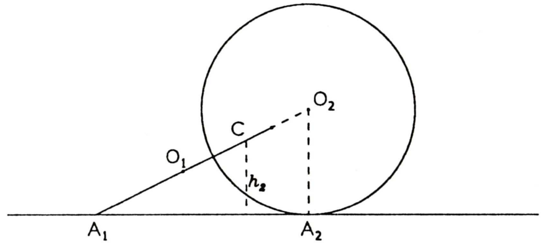


FIGURE 3

Thus $\eta_1 = \eta_2$ if and only if $(r+d)/(r+2d) = 1/\sqrt{2}$, which gives us $\delta = d/r = 1/\sqrt{2}$ or $2\delta = 2d/r = \sqrt{2}$ as the relative separation of the centres.

This shows that $2\delta = \sqrt{2}$ is necessary for the centre of gravity to have constant height, but this only considers the two special positions. The general position requires some care. As in Position 2, let the circles touch the base plane at A_1 and A_2. The intersection of the plane of disc 1 with the base plane is the tangent line to circle 1 at A_1. The tangent lines at A_1 and A_2 intersect at a point O in the base plane. The line of centres O_1O_2 also passes through O, which can be viewed as the intersection of the planes of the two discs and the base plane.

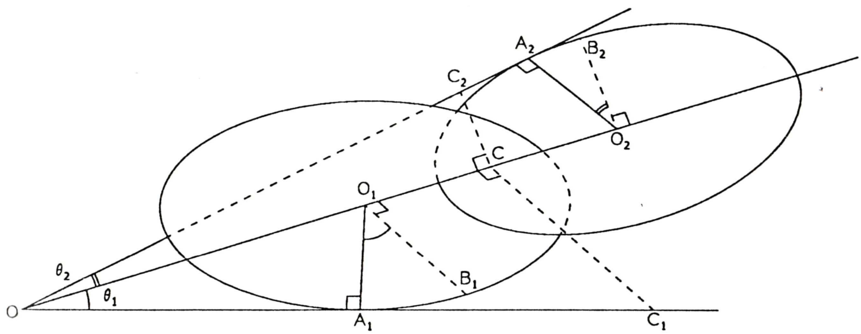


FIGURE 4

This is shown in Figure 4 where the line of centres OO_1O_2 is above the base plane which contains O, A_1 and A_2. Then OO_1A_1 and OO_2A_2 are right triangles. Let $\theta_i = \angle O_iOA_i$. Draw O_iB_i in the plane of disc i and perpendicular to OO_1O_2.

Then $\angle OO_iB_i = 90°$, hence $\angle A_iO_iB_i = \theta_i$, and so θ_i is the angular measure of A_i from the natural point B_i which is where the Wobbler touches the base plane in Position 1.

Now draw axes CC_1, CC_2 through C and parallel to O_1B_1 and O_2B_2. Then CO, CC_1, CC_2 are cartesian axes based at O. We have $OO_1 = r \operatorname{cosec} \theta_1$, and so $OO_2 = OO_1 + O_1O_2 = r \operatorname{cosec} \theta_1 + 2d$, which gives the basic relation:

$$\operatorname{cosec} \theta_2 = \operatorname{cosec} \theta_1 + 2\delta \tag{1}$$

[Finding this was the hardest part of our analysis.]

We have $CC_i = CO \tan\theta_i$, so the equation of the base plane with respect to the CO, CC_1, CC_2 coordinates is

$$x/CO + y/CO\tan\theta_1 + z/CO\tan\theta_2 = 1$$

Using the standard formula for the distance of a point to a plane, we have that the distance of C from the base plane, ie the height of the centre of gravity, is

$$\begin{aligned} h &= CO/\sqrt{1+\cot^2\theta_1+\cot^2\theta_2} \\ &= (r\operatorname{cosec}\theta_1 + d)/\sqrt{\cot^2\theta_1+\operatorname{cosec}^2\theta_2} \end{aligned}$$

so the relative height η is

$$\begin{aligned} \eta = h/r &= (\operatorname{cosec}\theta_1+\delta)/\sqrt{\cot^2\theta_1+(\operatorname{cosec}\theta_1+2\delta)^2} \\ &= (\operatorname{cosec}\theta_1+\delta)/\sqrt{2}\sqrt{\operatorname{cosec}^2\theta_1+2\delta\operatorname{cosec}\theta_1+2\delta^2-1/2} \end{aligned}$$

The expression in the second radical is

$$(\operatorname{cosec}\theta_1+\delta)^2+\delta^2-1/2$$

Hence the relative height η is constant at $1/\sqrt{2}$ if and only if $\delta^2 = 1/2$, i.e. $\delta = 1/\sqrt{2}$ or $2\delta = \sqrt{2}$.

This establishes our first result—the generalised Wobbler has centre of gravity at constant height if and only if $2\delta = \sqrt{2}$.

Physically, this implies that the Wobbler is (meta-)stable in any position. We have actually verified this by showing that the centre of gravity lies over the line A_1A_2, joining the two points of contact, if and only if $2\delta = \sqrt{2}$. The point C', which is in the base plane directly under C, has coordinates $r^2/2CO(1, \cot\theta_1, \cot\theta_2)$, while

$$\begin{aligned} A_1 &= (\ \ d + r\sin\theta_1, r\cos\theta_1, \quad 0 \quad) \qquad \text{and} \\ A_2 &= (-d + r\sin\theta_2, \quad 0, \quad r\cos\theta_2). \end{aligned}$$

Then $A_1 - A_2 = t(C' - A_2)$ holds if and only if $\delta = 1/\sqrt{2}$ and then $t = 2 + 2\delta\sin\theta_1$. [It has just struck us that the curve of C' might be interesting to look at.]

3. The distance between contacts.

In the last paragraph we gave the coordinates of the contact points A_1 and A_2 with respect to the cartesian system based at C. Let $L^2 = (A_1A_2)^2$ and $L/r = \lambda$. We have

$$\begin{aligned} \lambda^2 &= (2\delta+\sin\theta_1-\sin\theta_2)^2+\cos^2\theta_1+\cos^2\theta_2 \\ &= 2+4\delta^2+4\delta\sin\theta_1-4\delta\sin\theta_2-2\sin\theta_1\sin\theta_2 \end{aligned} \tag{2}$$

In Position 1, $\theta_1 = \theta_2 = 90°$, so $\lambda_1^2 = 2 + 4\delta^2$.

In Position 2, $\theta_1 = 90°$, $\sin\theta_2 = r/(r+2d) = 1/(1+2\delta)$, giving $\lambda_2^2 = 4\delta^2 + 4\delta$. Thus $\lambda_1 = \lambda_2$ if and only if $2\delta = 1$, i.e. the centres are separated by the radius, or each centre lies on the other circle. When $\delta = 1/2$, (1) yields $\sin\theta_1 = \sin\theta_2 + \sin\theta_1\sin\theta_2$. This shows that λ is constant if and only if $\delta = 1/2$. As we have seen, this does not give a constant height of centre of gravity and so this device will not roll with the surprising ease of the Wobbler. [1] shows a version by Paul Schatz rolling down a plane.

4. Other Problems

We are primarily interested in the Wobbler, i.e. the device with $2\delta = \sqrt{2}$, but many of the following problems can be considered for the generalised Wobbler.

We have studied the practical problems of the physical thickness of the discs and the effect of rounding the edges of the discs and have found adequate solutions for small thickness. [We let $2t$ be the thickness and $\tau = t/r$ be the relative thickness.] However, we have had no success with the following problems:

A. As the Wobbler rolls on the base plane, the contact points A_i trace curves which appear somewhat like cycloids, or perhaps epicycloids. The curve traced by the point O also seems interesting. What are these curves?

B. Observation shows that the centre of gravity C does *not* move in a straight line, but oscillates. What is this curve?

Problem B shows that Problem A is a bit harder than one initially thinks. The oscillation of C shows that there is some frictional effect between the Wobbler and the base plane. Obviously, if there were no friction, the Wobbler would slide along. It seems that we can assume the Wobbler rolls without slipping, but this gets out of our knowledge and we are not clear about this. If so, then at a given position of the Wobbler, as in Figure 4, the arc lengths of the curves can be taken as $s_i = r\theta_i$. One can find all the relevant angles and lengths in the diagram, but we cannot see how to relate them to coordinates in the base plane. If this can be done, then one can also consider the next problem.†

C. How does the axis of centres OO_1O_2 move as the Wobbler moves?

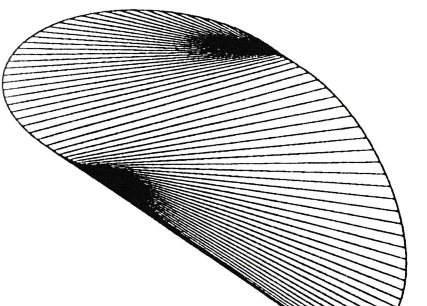


FIGURE 5

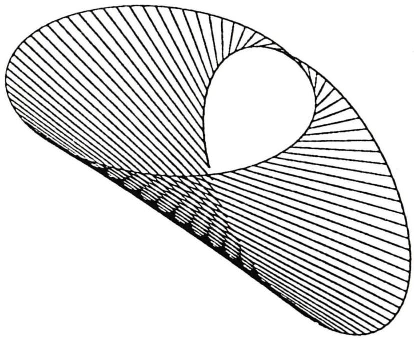


FIGURE 6

D. If we connect all the corresponding points A_1 and A_2, we enclose the convex hull K of the Wobbler, shown in Figure 5. What are the volume and the surface area of K? By looking at one eighth of K and estimating the volume inside, we find that the volume is at least $3.2r^3$. ‡

† John Watts has found some numerical results.

‡ John Watts has obtained $V = 3.2818r^3$ and $SA = 13.9236r^2$.

E. We can join the circles with lines to produce a doubly-twisted surface as in Figure 6. This should be the minimal surface generated by the two circles, i.e. the soap film, if done correctly. Can it be done correctly? If so, describe the surface and find its area.

REFERENCES

[1] Müller, George, ed. *Phänomena—Eine Dokumentation zur Ausstellung über Phänomene und Rätsel der Umwelt an der Seepromenade Zürichhorn, 12 Mai–4 November 1984*, Zürcher Forum, Zürich 1984, p. 79.

[2] Schatz, Paul. Swiss patent 500,000.

Clerihews

G. S. Murchiston

Herbert Grassman
Was a far from crass man.
He gave up mathematics, but made a
living translating the Rig-Veda.

A Director of Studies writes:

Linear analysis
Can cause complete paralysis
But we never weary
Of Galois theory

Pierre Deligne
Invented a machine
And just for a laugh
He called it SGA $4\frac{1}{2}$.†

Or, in French,

Pierre Deligne
Inventa une machine,
Et pour amuser ses amis
Il l'appellait SGA $4\frac{1}{2}$.

† SGA $4\frac{1}{2}$ should logically have come between SGA 4 and SGA 5 but was not finished until after the latter's publication, hence its name.

The Professor's New Book

Kevin Murphy

A Review of *The Emperor's New Mind* by Roger Penrose.

Can a computer have a mind? Roger Penrose, contrary to much currently popular scientific opinion, argues not. Before examining his reasons, let us first see why anyone would make such a claim in the first place.

It is a tacit assumption of most scientists that any physical object can be regarded as carrying out some computation. It is further assumed—by Penrose, too—that the mind is the result of the brain alone. Since the brain is undoubtedly a physical object, one might expect it to follow that the mind is a result of a series of computations (or algorithms) performed by the brain; that is, that the brain is a computer.

What exactly *is* a computer? It is defined as a machine which can calculate any *computable function.* A computable function is in turn defined as anything which can be performed by a mathematical abstraction called a Turing Machine (TM). This is called the Church-Turing thesis. Other definitions of computability, such as Church's Lambda Calculus, can be shown to be equivalent to the TM definition.

Penrose discusses algorithms and Turing Machines in some detail in his book, but it would not be relevant to go into that here. What is important is that I have not mentioned transistors or "silicon chips". What a computer is made of in no way affects what it can, in principle, do, provided it meets some very minimal level of complexity; Weizenbaum, for example, has shown how to make a computer out of a toilet roll and a pile of stones! That is, any hardware feature can always be emulated in software, a fact exploited every day in "real" computers. It follows that it should be possible, at least in principle, to "skim the algorithms off" the brain's neural substrate and re-implement them on, say, a digital computer, *and hence give a computer a mind.*

This view is called *Strong AI* (Artificial Intelligence) by John Searle, a philosopher at the University of California, Berkeley (the term *functionalism* is also used). Searle sums it up in the phrase, "The mind is to the brain what the program is to the hardware, and thus we can understand the mind without doing neurophysiology."

Searle is strongly opposed to strong AI. His criticism is based on his notorious "Chinese Room" thought experiment, in which he effectively asks you to imagine being a computer, and to "feel the lack of understanding". Picture yourself locked in a room. A series of Chinese symbols is passed into you; you are also given a set of formal rules—in English—which tell you what to do with these symbols, and what to pass back out. This is supposed to be analogous to a computer program. Imagine, further, that the symbols you pass out—"answers"—in response to the symbols passed in—"questions"—are indistinguishable from those of a native Chinese speaker: can we infer from this that you understand Chinese? Searle says no, and goes on to conclude that a computer could never understand anything, because it has nothing which the person in the room has not got. Penrose agrees.

Searle is left with a problem, however: what is it about the brain of the native Chinese person that lets him or her understand, and why can this not be emulated by a computer? He has no adequate answer, and is forced to utter such meaningless protestations as "computers don't have the right 'causal powers'", or, worse, computers are "not in the business of understanding." Supporters of strong AI accuse him of "wetware chauvinism". ("Wetware", by analogy with hard- and software, is the biological stuff brains are made of!)

This is where Penrose comes to the rescue. He claims that brains exploit some *non-algorithmic aspect of physics*, and *that* is why they cannot be emulated—or even fully simulated—by a computer (which, by definition, can only perform algorithms). "The kind of issue I am trying to raise is whether it is conceivable that a human brain can, by harnessing the power of appropriate 'non-computable' physical laws, do 'better', in some sense, than a Turing Machine" (p. 172). To paraphrase Searle, Penrose is claiming "We cannot understand the mind without doing *physics*." And, as we shall see, existing physics is inadequate for the job!

Another tacit assumption made by most scientists is that the brain can be *simulated*. Significantly, in his famous book *Gödel, Escher, Bach*, Douglas Hofstadter remarks (p. 572): "a computer simulation of a neural network is in principle feasible, no matter how complicated the network, provided that the behaviour of individual neurons can be described in terms of computations which a computer can carry out. This is a subtle postulate which few people even think of questioning." But it is precisely this that Penrose *is* questioning! Cf. the difficulty Searle has in explaining why a simulation of a Chinese person's brain would not understand Chinese!

Searle's argument, then, is one reason why Penrose is opposed to strong AI. However many people find this unconvincing. A stronger argument, first proposed by Lucas in 1961, is based on Gödel's Incompleteness Theorem (1931). This states that, within any given consistent formal system of sufficient complexity, there exist formally undecidable propositions.

A formal system is a mathematical system consisting of an alphabet of allowed characters, a set of axioms and a set of rules for deriving theorems from the axioms and previously derived theorems. A *consistent* formal system is effectively one which does not allow both P and $\sim P$ (the logical inverse of P) to be theorems. Gödel's Theorem says that any such formal system of sufficient complexity must be *incomplete*, that is, there must be a statement P which can be expressed in it but such that neither P nor $\sim P$ is a theorem of the system, although we of course expect exactly one of these two statements to be true (as it happens the construction is such that it is clear which statement we would consider to be true).†

What are the implications of Gödel's Theorem? For Lucas, it shows that the mind cannot be a formal system. This follows because we can always "out-Gödelise" any formal system. Can we not patch up our system by adding the unprovable statement as an axiom? No—because we can always pull the same trick on the new system! There is no limit, in principle, to how often we can do this: the system is said to be *essentially incomplete.*

We must be careful here, however. For while it can be *proved* that any formal system can, in principle, be Gödelised, this does not show that we ourselves can always do it. In fact, as there is no algorithm for Gödelising, it is rather unlikely that we could do it indefinitely. Hence there may exist a formal system of equal power to the mind. This is one argument used by Hofstadter against Lucas ([5] p. 475).

---

† In fact Gödel's Theorem also requires the axioms of the system to be *recursively enumerable*: that is, given any statement, is is possible to decide in finite time whether or not it is an axiom.

For Penrose, it shows "that the mental procedures whereby mathematicians arrive at their judgements of truth are not simply rooted in the procedures of some specific formal system." This is nothing controversial: after all, there is a result in mathematics known as Church's Theorem ([5] p./,579)—"there is no decision procedure for theoremhood"—which amounts to the same thing. What is controversial is the inference that Penrose makes, namely that there must be something fundamentally non-algorithmic about the brain to account for this. (The alternative strong AI approach would be to say: the brain is algorithmic, but this gives rise to a higher, informal level, which we call the 'mind'. This can, incidentally, be used as another argument against Lucas; See [5] p./,577.)

Furthermore, he argues that "the non-algorithmic forming of judgements is the hallmark of consciousness". (My own view is exactly the opposite, as we shall see.) He bases this view on his experiences as a mathematician. His argument seems to go: mathematics requires judgement about what is relevant, what is right, what is beautiful, etc; it also requires conscious thought; hence these judgements must be conscious. To me this conclusion seems a non-sequitur.

We have seen why Penrose believes physics must be non-computable. Is it true? Let us first ask what it means to say that a system is non-computable. Imagine we have a set of equations which completely describes the system (i.e. it is deterministic). We can see how it behaves by simply "plugging in" the right numbers. If the system is non-computable, there is no short cut algorithm for calculating the answers produced by this model of the system (Chaitin's definition of "randomness" as algorithmic incompressibility (see later) would seem to be relevant here, but Penrose does not pursue the issue). Note that computability implies determinism, but not vice versa.

The brain is a macroscopic object, and most neurophysiologists and physicists would agree that it can be adequately explained by classical physics. Is there non-computability in classical physics? Could this be exploited by the brain? There are at least two kinds of non-computability in classical physics: an in-principle kind and an in-practice kind. Neither would seem to be very "useful", however. Consider first the finding of Pour-El and Richards, that there exist certain kinds of computable data which, when "fed into" Maxwell's equations for electromagnetism, yield a non-computable result. This is not useful in practice because the data must be of a peculiar kind—not twice differentiable. Secondly, there is the much publicised problem of "sensitivity to initial conditions", or chaos. Again, this is surely not very exploitable.

Can a computer be made on the basis of classical physics alone? Penrose argues not. Consider, for example, a digital electronic computer. It relies on discrete electronic signals. Ultimately, these are based on the quantised nature of reality. What if we made a computer out of billiard balls bouncing off walls (as proposed by Fredkin and Toffoli)? That too implicitly relies on Quantum Theory, since the solidity of the walls cannot be adequately explained without it.

What about a "quantum computer", as proposed, for example, by David Deutsch? This exploits "quantum linear superposition" to perform calculations in parallel. Under certain contrived circumstances, it can outperform a Turing Machine in terms of the time taken to solve a problem: specifically, problems not in **P** can nevertheless be solved in polynomial time. However, it offers no in-principle advances, and is thus of little relevance. This is just as well, for it would be almost certain not to work in a brain, which is too "hot"—i.e. emits too much "noise"—to preserve quantum coherence.

Are there parts of the brain that work on a quantum level? Yes: the retina (which is an outgrowth of the brain) contains rods, which are sensitive to individual quanta of light (photons). However, this is the only known case of quantum-sensitive neurons. It is also significant that a human cannot perceive individual photons: they are deliberately suppressed to reduce visual noise. For similar reasons, I think it unlikely that there will be quantum-sensitive neurons in the brain proper: the system would not be robust enough.

Interestingly, Sir John Eccles, well known for his dualist interactionist views, has also speculated that quantum effects might play a rôle in the brain. He points out that if Heisenberg's Uncertainty Principle is applicable to objects the size of vesicles ($\sim$400nm), there would be uncertainty in their position of the same order of magnitude as the processes involved in normal neural communication. "All that would be required for an effective action of a mental event on a neuron would be an alteration of the probabilities of quantum release, which could be accomplished by a non-material agent such as a mental intention." ([4] p. 303). However, as Penrose points out (p. 359) "It is part of the theory that one *cannot* influence these probabilities: quantum theoretical probabilities are *stochastic.*"

There is in fact a deep mystery when we apply Quantum Mechanics (QM) to the brain, one which seems to have been largely overlooked (but see Morowitz). The "standard" interpretation of QM (the "Copenhagen Interpretation", after Niels Bohr, from Copenhagen, who was a dominant figure in the development, both mathematical and conceptual, of QM) is somewhat vague about when one should stop using the determinstic Schrödinger equation—which Penrose, after von Neumann, calls **U** (Unitary evolution)—and start talking in terms of probabilites—which Penrose calls **R** (Reduction of the state vector). Or, as it is sometimes put, when should one "collapse the wave function"? The standard answer is to say: when a measurement is made. But what is a measurement? This is called the *Quantum Measurement problem.* It is usual to refer to an "observer". So who observes your brain? Itself? God? One is reminded of Ronald Knox's limerick (in response to Bishop Berkeley's *"esse est percipi"*), and the anonymous response to it:

There once was a man who said, "God
Must find it exceedingly odd
    If he finds that this tree
    Continues to be
When there's no one about in the Quad."

Dear Sir, your astonishment's odd:
I am always about in the Quad.
    And that's why the tree
    Will continue to be,
Since observed by,
                Yours faithfully,
                        God.

Penrose believes that the **U**/**R** (quantum measurement) problem will not be solved until we have a quantum theory of gravity. Most physicists, including Penrose, believe that the rôle of quantum gravity is to eliminate the infinities that current theories produce at space time singularities (e.g. inside black holes). What has that got to do with the **U**/**R** problem? Penrose is proposing an involved argument, which goes roughly as follows. "All the successful theories of physics are symmetrical in time" (p. 302). That is, you can replace t with $-t$ ("run time backwards") and get the same answer. However, we perceive time as flowing forwards. This is usually explained in terms of the Second Law of Thermodynamics. But that rests on the assumption that the entropy at the Big Bang (or, in general, any initial singularity) was low, whereas the entropy in a final singularity (e.g. in a black hole, or the Big Crunch) should be high. This can be explained if the Weyl tensor (part of the Riemann tensor that describes the curvature of space-time) is 0 at initial singularities. Penrose calls this the Weyl Curvature Hypothesis (WCH). Some physicists regard this as a "God given boundary condition" which must just be accepted. Penrose finds this unsatisfactory, because "the restriction imposed on the Creator" by WCH is of the order $10^{10^{123}}$, which is far too great to go unexplained. Penrose believes that a *new* theory is required, which he calls CQG (Correct Quantum Gravity! by analogy with Quantum Electro Dynamics, etc.) in order to explain WCH and hence the origin of time asymmetry. He warns (p. 371) "It is my opinion that our present picture of reality, particularly in relation to the nature of time, is due for a grand shake up—even greater, perhaps, than that which has already been provided by present-day relativity and QM".

Traditional approaches to quantum gravity involve attempts at modifying general relativity. They thus implicitly treat QM as inviolate. Penrose argues exactly the reverse. (As Martin Gardner puts it in the foreword, "Penrose is one of an increasingly large band of physicists who think Einstein was not being stubborn or muddle-headed when he said his 'little finger' told him that QM is incomplete." Cf. p. 297 "I believe that one must strongly consider the possibility that QM is simply *wrong* when applied to macroscopic bodies".) He regards the **U/R** problem as a fundamental weakness of QM, which cannot be explained away by adopting a suitable "interpretation". He links this up with WCH by considering the "Hawking's Box" thought experiment. Briefly, his argument is this (p. 359ff.). Black holes "swallow up" matter and energy, and hence destroy possibilities ("flow lines merge in phase space"). This must (by Liouville's Theorem) be counterbalanced by a corresponding *creation of possibilities* ("flow lines must multi-furcate"). Penrose believes **R** is the cause of this. Thus **R** *is a quantum gravity effect!* Note that it is also time asymmetric.

When does the state vector reduce? When the curvature of space-time becomes "significant", i.e. reaches the one graviton level. "Note that, according to this idea, the procedure **R** occurs spontaneously in an entirely objective way, independent of human observation" (p. 368). This typifies Penrose's attempts at an objective interpretation of QM.

There is one remaining problem (pp. 290, 371). When the wave function "collapses", it does so instantaneously, no matter how "long" it may be. This *non-local* aspect of QM (as exemplified by the EPR paradox and Aspect's experiment) presents serious difficulties for special relativity. This too would have to be solved by CQG.

How does all this relate to the brain? Firstly, of course, Penrose's objective criterion for state vector reduction resolves the paradox of the brain "observing" itself. Then there is the apparent "oneness" of consciousness. By this he means the fact that we seem to be able to be conscious of only one thing at a time. "If a conscious 'mental state' is in some way akin to a quantum state, then some form of 'oneness' or globality of thought might seem more appropriate than would be the case for an ordinary parallel computer" (p. 399) (I shall present my own alternative later).

Penrose also uses the idea of "quantum consciousness" to resolve the "teletransportation paradox". Imagine a machine which can analyse your body into its constituent parts, and then transmit this information (at the speed of light) to a distant machine which can recreate you from scratch—as in *Star Trek*. (Note, as Penrose points out (p. 27), "there is no significance in preserving the identity of any particular atom. The question of the identity of any particular atom is not even meaningful." At what level *does* identity arise? That is another matter altogether ...) Imagine further that they forget to kill you! Thus there will be two copies of you—the "original" and the "new".

Penrose asks, "Is there anything in the laws of physics which could render a teleportation *in principle* impossible?" If awareness is a quantum state, then QM tells us that we cannot copy it without destroying the original, and thus the paradox is resolved. (This follows from the following line of reasoning (p. 270). Suppose we could copy an electron's spin state. If we could do it once, we could do it again and again. The resulting system would have a measurable angular momentum. This contradicts a fundamental result in QM which states that the spin cannot be measured.)

An alternative resolution of this paradox, which I personally prefer, involves the notion of algorithmic incompressibility. This does not rule out teleportation in principle, but renders it useless. The assumption is that there is no quicker way to build a human, given all the necessary information, than to simulate every stage of that person's life in real time. The resulting person will thus be different to the "original", because of the new experiences accumulated while waiting for the simulation. An alternative approach is to deny the possibility of extricating a person from his or her surroundings. People are more like beads on a spider's web than isolated billiard balls: see Parfit and Johnson, for example. Or, finally, one could deny that there is any paradox at all—Penrose certainly hasn't shown there is one.

Penrose also proposes a more "concrete" role for QM in the brain. To understand this, we must make a brief foray into the world of tiling patterns. There is a theorem which states that the only rotational symmetries possible for a crystalline pattern are twofold, threefold, fourfold and sixfold. Penrose discovered an aperiodic pattern that includes fivefold symmetry, and yet is "almost" a crystal. It is now called a quasi-crytsal. Furthermore, in 1984 such a quasi-crystal was actually discovered to exist in nature! The interesting question arises of how such a crystal is constructed. "The 'best' arrangement of the atoms cannot be discovered simply by adding on atoms one at a time in the hope that each individual atom can get away with solving its own minimising problem. Instead, we have a global problem to solve" (p. 437). He speculates that this growth is quantum mechanical in nature.

When a baby is born, its brain is not completely "hard-wired". There is a large amount of plasticity, which allows the environment to influence the developing brain. (Some astonishing work has been performed on the developing visual cortex of kittens. If they are brought up in an environment without, say, vertical lines, they will be incapable of "seeing" vertical lines in adult life. The neurons which are normally sensitive to lines of that orientation will simply not have developed.) But there is plasticity even in the adult brain, if only to allow memories to be stored and things to be learnt. Penrose discusses one particular variety of plasticity, in which the spines on dendrites can grow or contract, and thus affect the chance that the neuron will fire. He believes this process is similar to quasi-crystal growth. Vast numbers of alternatives would be tried out in parallel, and the global optimisation thereby achieved. "I am speculating that the action of conscious thinking is very much tied up with the resolving out of alternatives that were previously in linear superposition. This is all concerned with the unknown physics that governs the borderline between **U** and **R** and which, I claim, depends upon a yet-to-be-discovered theory of quantum gravity—CQG!" (p. 438).

My own opinion on these matters, for what it is worth, is this. The brain is a vastly complicated, non-algorithmic, parallel, distributed neural network. Human brains, at least, have evolved the remarkable ability to *emulate a Virtual Von Neumann Machine* (VVM), which we call consciousness. All of our analytical thought is carried out by this VVM.

This model has a number of attractive features. Firstly, the fact that we seem to be conscious of only one thing at at time (in contrast to the way we perform vast number of unconscious tasks in parallel) is explained by the sequential nature of a von Neumann machine. Secondly, the fact that we are conscious of what goes on in the VVM explains why we are fooled by our introspection into thinking that it is typical of our thought processes, whereas, in reality, it is only suitable for a very small subset of problems. Note, in particular, that this *virtual* machine is based on non-algorithmic, parallel hardware.

Let me apply my model to the question of mathematical creativity, which Penrose discusses at length. I would argue that the insights, the original, creative thoughts, are generated by the non-algorithmic, unconscious machinery of the brain. These are then "fed into" the VVM for rational scrutiny. The machinery responsible for creativity is far more biologically useful than that which lets us perform mathematics. Hence it is to be expected that it should exist in other animals. There is indeed evidence that non-human animals, at least among primates, can exhibit insight. Penrose mentions Lorentz's observation of a chimpanzee climbing on a box to get an overhanging banana. There are many other examples that I do not have time to go into (see [1] pp. 116, 131 for example).

I would argue that it is unnecessary to appeal to hypothesised non-algorithmic aspects of fundamental physics, provided it is the case that neural networks can outperform algorithms. I say this because neural networks—which Penrose only mentions once, and then only briefly—seem to provide everything Penrose is looking for, and are much easier to swallow (figuratively speaking). Consider the global optimisation problem referred to above. Neural networks excel at that (the most commonly cited example is the "Travelling Salesman" problem. It is significant that there is no known algorithm for this running in polynomial time). Consider Gödel's Incompleteness Theorem: that wouldn't apply to a neural network. Consider the Chinese Room argument: even Searle moderates his views when it comes to neural networks. Consider the difficulty of programming computers to do pattern recognition, of implementing "contents addressable memories", of formalising creativity: all of these (well, all except the last—I remain ever hopeful!) can be solved by neural networks.

I believe that neural networks are more than a fashionable trend, and present a genuinely new way of thinking about the brain. This rests on the assumption that they can outperform TMs. I justify this by referring to the relative ease with which neural networks can solve problems for which no algorithm is known. This does *not* imply that no such algorithm exists, of course. Since I accept that neurons are computable, I accept that an algorithm which simulates the brain is always, in principle, possible (I am also prepared to accept that a brain simulation would be conscious). But this is computability in an uninteresting sense: it is too sensitive to the *details.* For instance, re-connecting one neuron would require an entirely new algorithm. It is commonly accepted that AI should try to model higher level phenomena. What I am claiming is that these higher levels *cannot* be modelled by algorithms. The *only* kind of computability there is in the brain, in general, is the uninteresting kind.

I say "in general" because there may well be a class of problems, presumably those we ourselves solve on our VVMs, which are amenable to algorithmic solutions. Indeed, the history of AI seems to support this. All of the successful programs have been in domains which require conscious, analytical, sequential thought e.g. theorem proving or chess.

However, there is a limit to how far one can succeed with algorithms alone. While it is true that a computer can now beat a grandmaster at chess (as of 1988), the way in which it does it bears little resemblance to the way in which the human does it. In an engineering context, this is not a problem: we want something that *works*—possibly even improves on the "original"—and we don't care how it does it. For those interested in the *science* of the mind, however, this is not an adequate response. It is also probably true to say that until we understand how the mind works, we will have little chance of improving upon it. The computer method uses "Brute Force and Ignorance"; the human seems to use a very sophisticated form of pattern recognition, or "chunking", in which he simply *does not see* bad moves, just as an amateur does not see *illegal* moves ([5] p. 286); this almost certainly exploits the same (non-algorithmic) pattern recogniton hardware as is used in vision. The difference could not be made clearer than by comparing the number of positions looked ahead per move. "Hitech" can look ahead 20–30 million moves per position. Human grandmasters tend to look ahead 20–30. When grandmaster Richard Reti was asked how many moves ahead he considered, he replied "One: the right one!"

In conclusion, this is a fascinating book, which touches on a vast range of issues, as befits its all-embracing subtitle "Concerning Computers, Minds and the Laws of Physics". As Gardner says in his foreword, "Penrose takes you on a dazzling tour that covers such topics as complex numbers, Turing Machines, complexity theory, the bewildering paradoxes of QM, formal systems, Gödel undecidability, phase spaces, Hilbert spaces, black holes, white holes, Hawking radiation, entropy, the structure of the brain, and scores of other topics at the heart of current speculations." It is not an easy book, certainly not for the uninitiated; Gardner accurately describes it as a "marvellous book for informed laymen".

The Emperor's New Mind, by Roger Penrose, OUP hardback £20.

REFERENCES

[1] Blakemore, C., *Mechanics of the Mind*, CUP, 1977.

[2] Blakemore, C. and S. Greenfield, *Mindwaves. Thoughts on Intelligence, Identity and consciousness*, Basil Blackwell, 1987.

[3] Dennett, D. and D. R. Hofstadter, *The Mind's I. Fantasies and reflections on self and soul*, Penguin, 1981.

[4] Eccles, J., "Brain and Mind, Two or One?" in [2].

[5] Hofstadter, D. R., *Gödel, Escher, Bach: An Eternal Golden Braid. A Metaphorical Fugue on Minds and Machines in the Spirit of Lewis Carroll*, Penguin, 1979.

[6] Johnson, M.,"Did I begin?" *New Scientist* 9 Dec 1989.

[7] Morowitz, H., "Rediscovering the mind", *Pyschology Today* August 1980. Reprinted in [3].

[8] Parfit, D., "Divided minds and the nature of persons" in [2].

[9] Weizenbaum, J., *Computer power and human reason.* W. H. Freeman 1976.

Book Reviews

[See also Kevin Murphy's article "The Professor's New Book" in this issue for a review of "The Emperor's New Mind" by Roger Penrose.]

The Mathematics of Games

by **John D. Beasley**. Reviewed by Mark Wainwright

This eminently readable and well set-out book is the fifth in the OUP *Recreations in Mathematics* series, in whose fortunes it should mark an upward turn. Its author sets out to show that, to the clear-headed, a vast range of games are susceptible of a simple but illuminating and effective mathematical treatment. This he does with consummate success, on the way debunking much that is spurious in a number of revered and conventional techniques for forecasting and grading results.

It is rare that one comes across a book of which one can truthfully say that "there is a delight on every page"; this book marks such an occasion. Furthermore, although Beasley writes for the general public, he makes no hidden simplifying assumptions, nor does he produce results like rabbits from a hat; so, though comprehensible to an intelligent layman, *The Mathematics of Games* will be of equal interest to *Eureka* readers.

A few of the delights: some early chapters furnish an investigation into the nature of randomness, with applications to such important pastimes as the game of Snakes and Ladders; Chapter 4 presents a marvellous contrast between three ball-games; Chapter 5 grapples with the problem of measuring the skill of players and predicting results. Beasley deals with the fundamental weakness affecting any attempt to extrapolate from known data: a footnote draws the moral for economic forecasts.

Proceeding via games of bluff through the analysis of puzzles, the book presents a thoroughly amusing analysis of Nim in various forms, comprehending the Sprague-Grundy theorem and looking briefly at misère play. Eventually, coming past games of pure skill and automatic games, Beasley concludes by describing Turing's solution to the Halting Problem and finally Gödel's theorem. Though it could be argued that the latter is much too often written about already, Beasley has at least the decency to restrict his remarks about it to a brief and well-written final section, redeeming, I think, the inclusion of material whose place in a book of this sort is dubious. In any event I hope that I could not have recommended this book more strongly than I have done: for if I could have then I have failed in a clear duty.

The Mathematics of Games, by John D. Beasley, OUP hardback £14.95.

The Puzzling World of Polyhedral Dissections

by **Stewart T. Coffin**. Reviewed by Tim Auckland

The brightly coloured cover and curiously mixed title suggest that this book is just another of those "let's take a trivial piece of mathematics and write down to the general public" books. The author soon dispels this false impression. This book will appeal to all those with a good grasp of low dimensional geometry whether they be mathematician or artist. After all, a book which starts with Escher and ends with Bach can't be all bad.

The author runs a small cottage industry designing geometrical puzzles and hand-crafting them in wood. The reader is encouraged to make for himself the models examined, and the book gives sufficient details of construction and basic woodworking techniques to make this possible (more woodworking details are given in the book "Puzzle Craft" by the same author). However, it is typical of the author's style to make these details necessary as well as sufficient. He believes that there is as much enjoyment to be found in trying to design the models as there is in using them. For example, he writes:

> Thirty such identical sticks are altogether impossible to assemble. The various schemes for modifying some of the pieces to permit assembly are left for the reader to invent ...

Indeed he states that at least one of the puzzles in the book has no solution at all. There are no detailed plans, the structures being deduced from accurately drawn diagrams of pieces and completed puzzles and from implied symmetries. I agree in principle with the author's procedure, but it is very frustrating to those without access to decent woodworking facilities.

There are detailed investigations of variants of familiar themes such as tangrams, soma cubes and truncated octahedra (known amongst Archimedeans as splats). Some mathematical analyses are presented, together with some computational investigations, but the majority of the puzzles are merely described and interesting points noted, any further investigation being left to the reader.

The author has gone to some trouble to find out as much history as possible for each puzzle. In many cases this is not a great deal, as many of the puzzles are "obvious" and so have been independently invented in many different places, and some are poorly patented. Many have been published as "ancient" or "from the Orient", despite such puzzles having been in existence for less than two hundred years.

In conclusion, the book is full of beautiful diagrams and fascinating ideas. If I had a small saw bench and drill press I would have made half the models in the book before writing this. I think it is probably worth the recommended price of £17.50, though it is even better value to write a review of it and get a free copy from *Eureka*.

The Puzzling World of Polyhedral Dissections,
by Stewart T. Coffin, OUP hardback £17.50.

Makers of Mathematics

by **Stuart Hollingdale**. Reviewed by John Aspden

Mathematics as presented to us undergraduates is a beautiful and precise subject of well-formed ideas, where rigour and definition of fundamental concepts are prime objectives. It was not always so. Dr Hollingdale has collected a number of short biographies of his favourite mathematicians, tracing the development of the field from ancient times to about the beginning of the twentieth century, where the ongoing research becomes tricky for the laymen at whom the book is aimed. The treatment is idiosyncratic and enlightening, spread throughout with snippets of ancient thought presented in such a way as to give the modern reader both a chance to understand the arguments and a feeling for the peculiar difficulties facing our forebears. Whilst the book is not going to give anyone any great insights into maths. itself, it does put the various ideas into their historical contexts. It's a fairly riveting read, full of gems like Euler's cheerful method of summing various infinte series by factorising $\sin(x)$ as an infinite product, and the tactful remark "opinions differ as to how much of this [Gauss's] information was passed on to the young Lobachevski".

The book's style varies from the iconic portrayal of the various legendary ancients through to coverage of the fierce in-fighting of more recent times: Cantor and Poincaré, Liebnitz and Newton; Galois gets to fight his duel again. The nineteenth century is dealt with sparingly because of the sheer number of deserving candidates, and the twentieth is represented only by Einstein because of its incomprehensibility.

There's a fair bit of actual maths. in here as well, but, reflecting the actual methods used, it's mostly in the straight edge, compasses and caffeine tablets mould favoured by the Greeks. On the other hand, I've now got a semi-inkling what general relativity is all about, and slightly more respect for principles of least action and the like.

I'm not sure why I liked this book as much as I did, but I certainly enjoyed reading it and would recommend anyone else to. I'm quite sure that anyone at all interested in his subject (although after the Cambridge experience there many be precious few of those ...) will find something amusing or intriguing in here.

Makers of Mathematics, by Stuart Hollingdale, Pelican paperback £7.99.

More Call My Bluff

Stephen Turner

From page 21

The correct answers are as follows:

1.	**Bang-Bang type**	(b)
2.	**Quadratix**	(a)
3.	**Contorted Fractions**	(a)
4.	**Syrtis**	(c)
5.	**Suanching-Shihshu**	(c)
6.	**Bornological**	(a)
7.	**Doetsch's Three-Line Theorem**	(b)
8.	**Eikonal**	(a)
9.	**Who said ...?**	(b)

In number **9** the complete quotation is: "Mathematics, rightly viewed, possesses not only truth, but supreme beauty—a beauty cold and austere, like that of sculpture."

The Inter-University Call My Bluff competition was held in Trinity OCR on the afternoon of Sunday 3rd December 1989. Teams took part from Bristol, Cambridge, Imperial College London and Nottingham. A team from King's College London arrived in Cambridge too late and couldn't find us, so spent the day exploring the town! The President, Frances Hinden, was in the chair, aided by myself as scorer, with Colin Bell representing the press (all right—the Archimedeans' Chronicles). Two points were awarded for guessing the correct definition, and one to the defining team for each team guessing a wrong definition. Apparently the Cambridge team was misbehaving for it twice had points deducted: 20 points for defining a non-mathematical word (at round 4) and, right at the end, a further 18 points for time-wasting. After seven rounds of fierce competition, lasting two and a half hours, the final scores were thus as follows:

Cambridge	0
Nottingham	18
Bristol	23
Imperial	43

So Imperial College were the clear winners, the prize for best false definition going to Nottingham. The contestants then invaded a local pizza establishment and dined there before scattering once more to the four winds whence they had come. It remains only to thank all who helped and all who took part from Cambridge and elsewhere.

Solutions to Problems Drive

Frazer Jarvis and Graham Nelson

From page 28

1.

[1] 6	[2] 3	[3] 9	[4] 2	[5] 1
[6] 5	4	6	[7] 6	8
[8] 6	5	8	[9] 9	0
[10] 7	[11] 8	[12] 1	6	[13] 5
[14] 2	4	0	1	9

(Start at the bottom left and it drops out).

2. 2, 10 and 15 are safe. The others may cause death (consider the possible programs).

3. 2 possibilities. ($1+2+3+4+5+6$ is the least total gain of places, 3×7 is the most that can be gained if there are three new entries ($10-3=7$). Then check cases.)

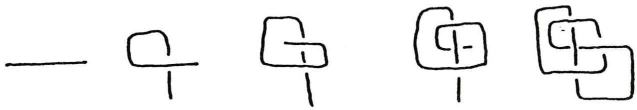
4. 3071 pages. (There are 240 days; add an imaginary "recovery" day since you may as well work on the day before the exam, then see which "chunks" of $n-1$ days work in a row plus 1 day off, benefit most. As it happens, the best way to do it is 58 chunks of 4, and 3 chunks of 3.)

5. 1/27 and 1/9. (Sphere touches big tetrahedron at the centres of the faces, 1/3 of the way down each altitude. Rotate small tetrahedron so that its vertices are at these points—clearly this then has height 1/3 of that of the big tetrahedron, giving 1/27 of the volume. For the cube, note that we can find 4 vertices which define a regular tetrahedron, thus having volume 1/27, but it's 1/3 of the volume of its containing cube by the hint, so that then has volume 1/9.)

6. $A=Y$; $B=W$; $C=Z$; $D=X$; $E=V$. (The duck is obvious. A must speak in reply to the first question, and "You!" is a lie and B can't say it. A can't say "Quack!" since it's unhelpful confusing himself with the duck. Then C is the person who replies "Quack!" to the first question, this being neither true nor ingratiating to the questioner. So X is the manic person. Incidentally, the first word to come into W's head was verified experimentally by the compilers.)

7. 84 ("Try lots of cases"—AFJ)

8.



You can just work round the knot in fact. (Not a very good question, but we think the game is fun. Which knots can you get?)

9. 17/64. (The countries NW and N of R are independently democratic with prob. 1/2 in October; this gives 4 cases. Checking we see that Gemland is reunified in November with probability 17/32 and then Grand Burgundy has (1/2).(17/32) = 17/64 chance of falling in December. January 1st is too late!)

10. A=k; B=h; C=j; D=b; E=e; F=i; G=g; H=d; I=a; J=c; K=f.
Mascheroni's constant is another name for Euler's constant, B.

11. (a) 71. (Fibonacci starting with 3 and 7.)
(b) 45. (If you start at 1, differences are exactly those numbers out of 1, 2, 3, ... which are not in the sequence already, so 2, 4, 5, 6, 8, ...)
(c) 43. (Alternate prime numbers.)
(d) (A protest against all previous series questions. Marks were awarded for imagination, ingenuity and wit (if any).)

12. 4560. (That's the least possible distance moved, so it's at least that many; look at smaller cases and it's clear you can attain it.)

The traditional prize of a bottle of port was awarded to Michael Earnshaw and Tim Wilkins, both of Peterhouse, after a close-run contest with Alex Selby and Alan Stacey. The former pair will set next year's questions.

The wooden spoon was awarded to *W* and Prof. St John (see questions **6** and **2** respectively).

Solution to "A Logical Problem"

From page 68

E is the Junior Treasurer. One key to the solution is that the first and last statements are the same. Assuming both are false, we know immediately that both *D*'s statements are true, so that *A* is the Junior Treasurer. This makes *C*'s first statement false, so *A*'s second statement must be true. Unfortunately, we now have a second Junior Treasurer. Thus our initial assumption is false. So we know that *D* is the Vice-President. A similar chain of reasoning allows us to identify the remaining committee members fairly easily.

The Cover

Thomas Bending

The design on the cover was inspired by The Archimedeans' symbol—a projection on to Euclidean 2-space of a sphere inscribed in a right circular cylinder of the same height as the sphere's diameter—and depicts a reflecting sphere inscribed in a clear, refracting cylinder with black ends, floating above a large planet covered with a chequerboard. It was produced as a bitmap by ray-tracing on the University IBM mainframe and converted to PostScript for printing on an Apple LaserWriter.

I would like to thank Ian Redfern for help with the PostScript, Mark Wainwright and others for much constructive criticism ("No, Tom, it still doesn't look right") and Robert Hunt for slick vector-based solutions to various geometrical problems, which I had forgotten after Part IB.