Eureka Editor
archim-eureka@srcf.net
The Archimedeans
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WA
United Kingdom

Published by The Archimedeans, the mathematics student society of the University of Cambridge

Thanks to the Betty & Gordon Moore Library, Cambridge

EUREKA

The Journal of the Archimedeans

Number 49 March 1989

Eureka 49

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.

Archimedeans' Committee 1988-1989

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 to you 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 this issue and possibly future ones 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:

The Arts School, Bene't Street, Cambridge CB2 3PY, England

Electronic mail from the U.K. to:

`archim@uk.ac.cambridge.phoenix`

and from the rest of the World, to:

`archim@phoenix.cambridge.ac.uk`

Eureka 49

Contents

Credits

Editor . . . . . **Graham Nelson**, Selwyn College
Business Manager . . . . . **Frances Hinden**, Trinity College

Number 49 March 1989

 ISSN Number 0071-2248.

Editorial

Fifty Years of Eureka

In January 1939, the first editors of Eureka, F. J. Collinson (Newnham), E. P. Hicks (St. John's) and A. Jackson (Emmanuel), set out in the first editorial the reasons for its existence. Quoting from the then President, they said:

> "... The chief thing is to make it interesting to every Cambridge mathematician, to help build up the corporate interest in the subject... to link together students, researchers and dons, other English and foreign universities. We must aim to stimulate informed discussion, especially as to Cambridge questions..."

They continue: "We realise how inadequately this first issue fulfils these high ideals." In fact, this is modesty on their part; they had received encouragement from the University of Chicago and the University of Wisconsin, for example, and their first issue is extremely thought-provoking. There are articles on mathematics as practised in other universities and countries, an article celebrating the 150th anniversary of the publication of Lagrange's *Méchanique Analytique*, a classic recreational article on Mathematics and Chess (beginning "Chess is a game in which 2^5 particles move in a finite space. In this respect it resembles the Universe, in which we are told there are $136 \cdot 2^{256}$ particles"), a history of the Mathematical Association, commentary on the beginnings of student representation in the running of the Faculty, and many other worthy items.

Looking through the very early Eurekas, one is struck by the fact that at once the times in which they lived were so different, and that the people making up the Society were so similar. In the war-time Eurekas, although little explicit reference is made, one finds (in issue 7) a Heffer's advertisement offering catalogues "as often as the paper control regulations allow"; it is recorded (in issue 3) that Maths teachers evacuated to teach evacuees were made members of the Society; collections were made for medical aid to the Soviet Union; and so on. There is debate as to whether Part IB students should be called up immediately, or be allowed to finish their degrees first. On a lighter note, the age of Eureka can be judged from such things as discussion of whether Noel Coward is considered "too lowbrow" by Cambridge mathematicians, and a book review (in issue 1) of E. T. Bell's classic *Men of Mathematics.*

In 1949, D. J. Wheeler (now Prof. D. J. Wheeler!) describes an amazing machine called EDSAC, being built by Dr. M. V. Wilkes (now Prof. M. V. Wilkes!) in which numbers are "stored in the form of supersonic bursts of waves travelling in mercury contained in a tube". This machine seems to have captured the interest of the Society, and much is written about its earlier years: although the modern reader may not think much of its abilities, he would do well to read J. C. P. Miller's report in 1951 (issue 14) that the record for the largest known prime (previously $2^{127}-1$, found by Lucas seventy-five years before) had been broken, using EDSAC to reduce the time needed for a Fermat test to base 2 to only a few minutes and had found

$$P = 5{,}210,644,015,679,228,794,060,694,325,390,955,853,335,898,483,908,056,$$
$$458,352,183,851,018,372,555,735,221.$$

On the other hand, the recreational content of Eureka has changed hardly at all, and the regular content of surreptitious or merely ludicrous content has always been with us.

There are far too many gems in Eureka to report them all, but reference must be made to the classic article in issue 2 by P. M. Grundy on "Mathematics and Games", which seems to be the first indication that the game Nim contains the theory of all such games. Other recreations have included finding which real numbers could be closely approximated using normal mathematical symbols but only four 4's (J. H. Conway and M. J. T. Guy, in issue 25, following up a puzzle in issue 13: it is rather hard to explicitly do for interesting reals such as Euler's constant γ, for example) and analyses of the Boat Race (issue 46), First Division football results (issue 33), collecting every different free gift from packets of breakfast cereal in a promotion (issue 36), and maximizing Tripos performance (issue 34). More recently, investigations appear of the possible proportions of pairs of elements of finite groups which commute (issue 43), and of the justly celebrated Audioactive Chemistry (issue 46).

Many people who have now become famous mathematicians of the past wrote for Eureka: G. H. Hardy, P. Hall, P. A. M. Dirac, W. Hodge and M. Cartwright to name but a few. The list of mathematicians in the Faculty here and elsewhere who wrote for or were connected with Eureka is longer yet: a startling proportion of the present lecturers and Professors here may be found in the pages of suitably early issues. In view of the difficulty in recruiting Business Managers for Eureka, I am tempted to observe that the present Head of D.A.M.T.P. is Professor H. K. Moffatt, and that one H. K. Moffatt was Business Manager in 1957...

There have been many humorous items in Eureka, but my favourite must be the classic article in Eureka 16 on "A Contribution to the Mathematical Theory of Big Game Hunting" by H. Pétard, which also (more famously) appeared in the *American Mathematical Monthly.* This classic monograph in the field draws results from such sources as the Trivial Club of St. John's College (now euphemistically known as the Adams' Society, some might think), to the problem of catching the lion, *felis leo* in the desert. Many methods are given, including:

The Peano Method Construct, by standard methods, a continuous curve passing through every point of the desert. It has been remarked that it is possible to traverse such a curve in an arbitrarily short time. Armed with a spear, we traverse the curve in a time shorter than that in which a lion can move his own length. □

The Schrödinger Method At any given moment there is a positive probability that there is a lion in the cage. Sit down and wait. □

A Topological Method We observe that the lion has at least the connectivity of a torus. We transport the desert into four-space. It is then possible to carry out such a deformation that the lion can be returned to three-space in a knotted condition. He is then helpless. □

Eureka has featured satires and parodies of just about everything, from the Ring cycle (issue 30) to two different mathematical versions of *1066 and All That* (issues 12 and 21). For many years it was customary to write poetry rather than prose, and a sample of the results may be found elsewhere in this issue. Another diverting item (in issue 39) is a collection of photographs of some of the lecturers of the time. Many of those depicted are still at Cambridge, and some have even changed their hair-styles. I particularly recommend the photograph of Dr. Körner.

I have been asked to reprint the details of the Archimedeans' tie, which was adopted in 1950 (as reported in issue 13): "the design consists of Archimedean spirals with $\epsilon\upsilon\rho\eta\kappa\alpha$ between them".

Since 1942, Eureka has almost always featured a Society column, in which our activities are recorded. This has ranged from thanking the speakers (who have included such worthies as Sir Arthur Eddington and Dr. F. Hoyle, along with just about everyone else who has been a mathematician here in the last fifty years) to veiled remarks about members falling in the Cam on the annual Punt Trip (which seems to have its origins in the antiquity of the Society). The most unlikely I can find reference to is the Telepathy Evening, reported in issue 13, at which the Archimedeans present were revealed to be (slightly) significantly worse at guessing Zener card designs than a random strategy. Reading through the Society columns, one observes a certain periodicity in their contents, but it is nevertheless amusing to note, for example, that Chris Zeeman threw the Archimedeans' Christmas Party in 1947, exactly forty years before the committee I was on did.

Similarly, the Problems Drive problems have been published almost every year since 1949, with the custom of this year's winners setting next year's problems apparently having survived since then. In addition, Eureka has always contained short problems along with the larger items; this issue is no exception. These range from such as: "Show that there exist intelligible sentences containing $(14 \cdot 3^n - 3)$ successive *had*'s, where n is any non-negative integer" (issue 18), to alphametics such as $\mathtt{EUREKA} + \mathtt{EUREKA} + \mathtt{AN} = \mathtt{ANSWER}$, contributed by Dr. Partington (in issue 38). Crosswords and crossnumbers have been particularly fruitful. Quite apart from an analysis of "crosswordiness" (issue 5), crosswords of almost every kind have appeared: with letters, with numbers, three-dimensional, hexagonal and even in Roman numerals ("Crux Verborum", in issue 12). The Eureka crossword makes a hopefully welcome return in this issue.

In conclusion, I can only say that I hope the first editors would approve of the course Eureka has taken through the years. May it continue for many more.

Acknowledgements

Firstly, I should like to thank the contributors for their time and effort, without whom Eureka could not exist.

Eureka 49 is the first to be produced (almost) entirely in TEX; which has been something of an adventure for the Editor. Far too many people have helped me with the typesetting to name them all here, but I am particularly grateful to Adam Chalcraft, Imre Leader, Matthew Richards, Dr. Jonathan Partington, Bob Dowling, George Russell and John Croft.

I should like to thank the Department of Applied Mathematics and Theoretical Physics, and in particular Dr. Mark Manning, for helping with the production of laser-printed copy.

Above all, I would like to take this opportunity to thank Frances Hinden, who has put an enormous amount of work into sorting out the Eureka finances and subscriptions. The Society is in her debt.

The Society

Gerard F. Thompson (Chronicler)

The Triennial Dinner of 1988, with Professor Christopher Zeeman and Dr Lickorish as guests of honour, was a splendid occasion. In particular, Professor Zeeman's practical demonstration of a boomerang during his speech enlivened the evening. However, this event was organised by the previous committee. Could the present committee live up to the record of last year's?

Fortunately, the Easter term is traditionally a quiet term for the Society, and so the committee had time to settle in before plunging into May Week. The social events were in general as successful as the weather permitted; the Garden Party was dry if not bright, but numbers for the punt trip were diminished by the chill wind and threatening sky, and only a small party of gallant Archimedeans eventually arrived in Grantchester. In the croquet match, the faculty exhibited unexpected facility, and won comfortably. However, the Dampers failed to take up the challenge to a punt joust, rumour having it that they were not in the peak of physical condition on the appointed day.

The evening speaker meetings were better attended than last year; the Society would like to thank Professor W. W. Sawyer, Dr. J. Cofman, Dr. I. G. Porteous, Dr. J. Rae, Professor Brown and Professor I. Alexsander for their interesting and stimulating talks. Mr. Winter of the University Careers Service organised another successful Careers Evening, and we are grateful to him for doing so. There are, at the time of writing, still four major speakers to come. The Society also thanks those of last year's speakers who came after the last Eureka was published: Dr. J. Butterfield, Professor R. Schwarzenberger and Professor Sir Michael Atiyah.

As regards the other serious activities undertaken by the Society, an Alternative Guide for first year mathematicians was again produced, providing information valuable in the first few weeks in Cambridge. The squash rivalled the incredibly successful one of last year, and over the whole year the Society gained more than 170 new members. This fact, and the proverbial prudency of the Junior Treasurer, have meant that the Society is still in a healthy financial state. The bookshop is running as smoothly as ever, and the problems with the (non-) distribution of Eureka over the past few years are being dealt with by the Business Manager.

The subgroups have enjoyed mixed fortunes over the year: while the Musical Appreciation subgroup has ceased to be, at least until someone volunteers to run it, the others have gone from strength.

So, as the present committee hurtles into oblivion, we pause to wish those about to be elected to the new committee the best of luck in the coming year.

Forthcoming events include a Scavenger Hunt (with or without the Invariants) and a version of "Just a Minute".

Unfortunately the avuncle of the Society, Mark Owen, has not contributed a quotation to sum up this year, so I will merely say that "The Archimedeans had another good year."

On Maximally Dense Forests

Matthew Richards

"Lucky we know the Forest so well, or we might get lost," said Rabbit half an hour later, and he gave the careless laugh which you give when you know the Forest so well that you can't get lost. – A. A. Milne, *The House at Pooh Corner*

How dense a subset of $\mathbb{Z}^n$ (n a positive integer) is it possible to find with the property that if we join all pairs of points in it which are nearest neighbours in $\mathbb{Z}^n$ we get a forest, i.e. a graph with no cycles?

Such was the problem of which the author became aware at the Puzzles and Games Ring of Sunday 17th April, 1988. In the case $n = 2$ it was clear that the subset could not contain all the four points of any little square of nearest neighbours, so that in some sense $\frac{3}{4}$ was an upper bound. (Similar arguments produced upper bounds for other n, e.g. the value $\frac{5}{8}$ when $n = 3$.) However, the best example known was the subset $\mathfrak{D}^2$ of $\mathbb{Z}^2$ consisting of two out of every three diagonals, the central portion of which is shown in Figure 1, having in some sense a density of $\frac{2}{3}$.

Subsequent work by many members of the Ring resulted in a general argument which proved that the bound $\frac{n}{2n-1}$ is in fact best possible, and this will form the first part of this article. We shall then investigate the properties of forests with this maximum density.

Let us begin by making precise the above notions. For $n \in \mathbb{Z}^+$, we define the Euclidean metric d on $\mathbb{Z}^n$ by

$$\mathrm{d}(\underline{u}, \underline{v}) = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2},$$

where $\underline{u} = (u_1, \ldots, u_n)$ and $\underline{v} = (v_1, \ldots, v_n)$ are elements of $\mathbb{Z}^n$. Then with each subset $\mathfrak{V}$ of $\mathbb{Z}^n$ we associate the (possibly infinite) graph $(\mathfrak{V}, E)$, where

$$E = \{\{\underline{u}, \underline{v}\} \mid \underline{u}, \underline{v} \in \mathfrak{V}, \mathrm{d}(\underline{u}, \underline{v}) = 1\}.$$

It should cause no confusion if we also use $\mathfrak{V}$ to mean $(\mathfrak{V}, E)$; in particular $|\mathfrak{V}|$ will mean the order of both $\mathfrak{V}$ and $(\mathfrak{V}, E)$. So we are interested in the subsets $\mathfrak{V}$ of $\mathbb{Z}^n$ where $\mathfrak{V}$ is a forest.

For $\underline{u}, \underline{v} \in \mathbb{Z}^n$, let $\underline{u} + \underline{v} = (u_1 + v_1, \ldots, u_n + v_n)$. This makes $\mathbb{Z}^n$ into an abelian group with identity $\underline{0} = (0, \ldots, 0)$ and inverses $-\underline{v} = (-v_1, \ldots, -v_n)$. Now, for $\mathfrak{V} \subseteq \mathbb{Z}^n$ and $\underline{u} \in \mathbb{Z}^n$, let

$$\mathfrak{V} + \underline{u} = \{\underline{v} + \underline{u} \mid \underline{v} \in \mathfrak{V}\}$$

and $\mathfrak{V} - \underline{u} = \mathfrak{V} + (-\underline{u})$. Then $\underline{u} \to (\mathfrak{V} \to \mathfrak{V} + \underline{u})$ is an action of $\mathbb{Z}^n$ on $\mathrm{P}(\mathbb{Z}^n)$. We say $\mathfrak{W} \subseteq \mathbb{Z}^n$ is a translate of $\mathfrak{V}$, and write $\mathfrak{W} \sim \mathfrak{V}$, if $\mathfrak{W}$ lies in the same orbit as $\mathfrak{V}$ under this action, i.e. if $\mathfrak{W} = \mathfrak{V} + \underline{u}$ for some $\underline{u} \in \mathbb{Z}^n$. Thus $\sim$ is an equivalence relation on $\mathrm{P}(\mathbb{Z}^n)$.

We shall only be concerned with those properties of $\mathfrak{V} \subseteq \mathbb{Z}^n$ which are invariant under translation, i.e. which are shared by all $\mathfrak{W} \subseteq \mathbb{Z}^n$ with $\mathfrak{W} \sim \mathfrak{V}$, for example the property of being a forest.

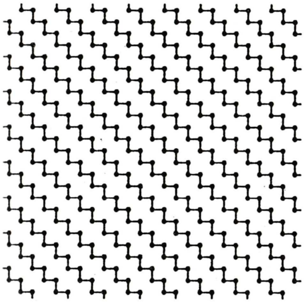


Figure 1 $\mathfrak{D}^2$

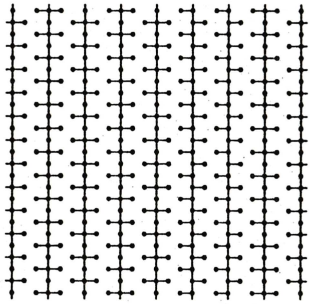


Figure 2 $\mathfrak{I}^2$

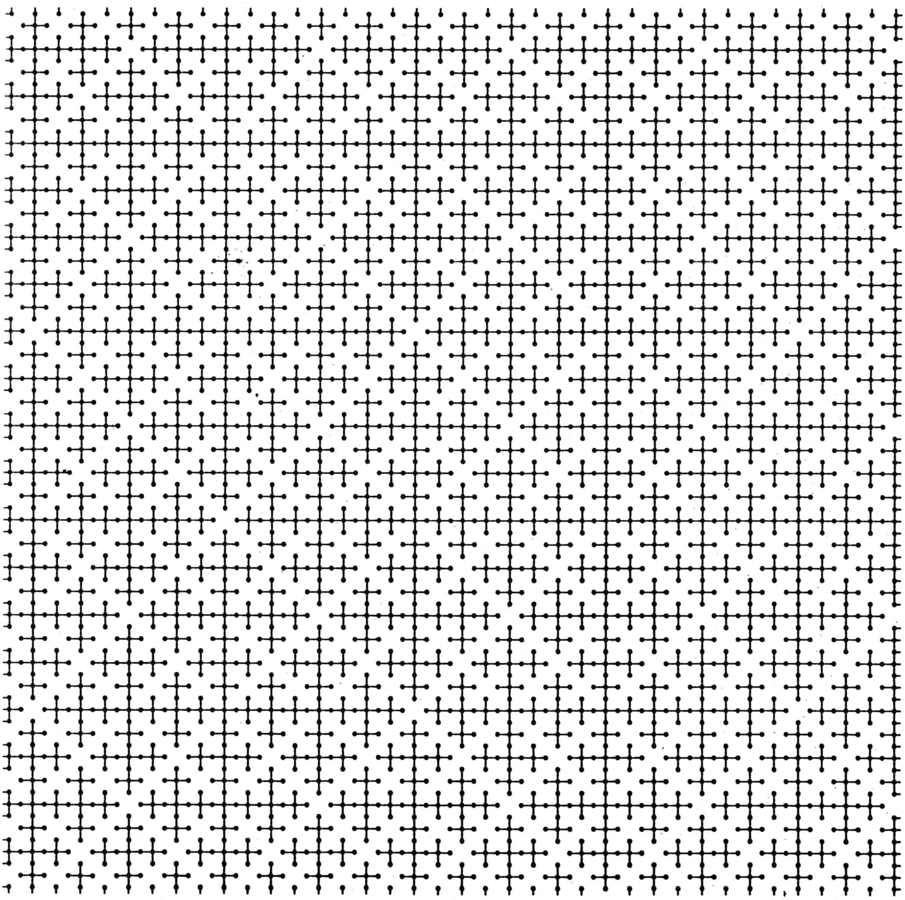


Figure 3 $\mathfrak{G}^2$

Now for each $\rho \in \mathbf{R}^+$, let $\mathfrak{B}^n_\rho = \{\underline{v} \in \mathbf{Z}^n \mid \mathrm{d}(\underline{v}, \underline{0}) < \rho\}$, and for $\mathfrak{V} \subseteq \mathbf{Z}^n$, let

$$\sigma_\rho(\mathfrak{V}) = \frac{|\mathfrak{B}^n_\rho \cap \mathfrak{V}|}{|\mathfrak{B}^n_\rho|}.$$

Then $0 \leq \sigma_\rho(\mathfrak{V}) \leq 1$, so we can define the upper density of $\mathfrak{V}$, $\sigma^+(\mathfrak{V}) = \overline{\lim}_{\rho\to\infty}\sigma_\rho(\mathfrak{V})$, and the lower density of $\mathfrak{V}$, $\sigma^-(\mathfrak{V}) = \underline{\lim}_{\rho\to\infty}\sigma_\rho(\mathfrak{V})$, so that $0 \leq \sigma^-(\mathfrak{V}) \leq \sigma^+(\mathfrak{V}) \leq 1$. If $\sigma^-(\mathfrak{V}) = \sigma^+(\mathfrak{V})$, we say $\mathfrak{V}$ has density their common value, which we denote $\sigma(\mathfrak{V})$.

This definition agrees with the ideas expressed above. In particular, $\sigma(\emptyset) = 0$ and $\sigma(\mathbf{Z}^n) = 1$. Note, however, that not all subsets have a density. We wish to find, for each $n \in \mathbf{Z}^+$,

$$\mu_n = \sup\{\sigma(\mathfrak{V}) \mid \mathfrak{V} \subseteq \mathbf{Z}^n, \mathfrak{V} \text{ a forest}, \sigma^-(\mathfrak{V}) = \sigma^+(\mathfrak{V})\}.$$

First let us check these definitions are translationally invariant. We need:

Lemma 1 *If $\mathfrak{V}, \mathfrak{W} \subseteq \mathbf{Z}^n$ and $\mathfrak{V} \sim \mathfrak{W}$ then $\sigma_\rho(\mathfrak{V}) - \sigma_\rho(\mathfrak{W}) = \mathrm{O}(\rho^{-1})$, so that $\sigma^+(\mathfrak{V}) = \sigma^+(\mathfrak{W})$ and $\sigma^-(\mathfrak{V}) = \sigma^-(\mathfrak{W})$. Also if $\mathfrak{V}$ has a density, so does $\mathfrak{W}$, and $\sigma(\mathfrak{V}) = \sigma(\mathfrak{W})$.*

Proof We have $\mathfrak{W} = \mathfrak{V} + \underline{u}$ for some $\underline{u} \in \mathbf{Z}^n$, so

$$\begin{aligned}\sigma_\rho(\mathfrak{V}) - \sigma_\rho(\mathfrak{W}) &= \frac{|\mathfrak{B}^n_\rho \cap \mathfrak{V}| - |\mathfrak{B}^n_\rho \cap (\mathfrak{V} + \underline{u})|}{|\mathfrak{B}^n_\rho|} \\ &= \frac{|\mathfrak{B}^n_\rho \cap \mathfrak{V}| - |(\mathfrak{B}^n_\rho - \underline{u}) \cap \mathfrak{V}|}{|\mathfrak{B}^n_\rho|} \\ &\leq \frac{|\mathfrak{B}^n_\rho \setminus (\mathfrak{B}^n_\rho - \underline{u})|}{|\mathfrak{B}^n_\rho|} = \mathrm{O}(\rho^{-1}).\end{aligned}$$

Similarly $\sigma_\rho(\mathfrak{W}) - \sigma_\rho(\mathfrak{V}) \leq \mathrm{O}(\rho^{-1})$, so $\sigma_\rho(\mathfrak{V}) - \sigma_\rho(\mathfrak{W}) = \mathrm{O}(\rho^{-1})$. Now

$$\overline{\lim_{\rho\to\infty}}\, \sigma_\rho(\mathfrak{V}) = \overline{\lim_{\rho\to\infty}}\, \sigma_\rho(\mathfrak{W}) + \lim_{\rho\to\infty}(\sigma_\rho(\mathfrak{V}) - \sigma_\rho(\mathfrak{W}))$$

so $\sigma^+(\mathfrak{V}) = \sigma^+(\mathfrak{W}) + 0 = \sigma^+(\mathfrak{W})$. Similarly $\sigma^-(\mathfrak{V}) = \sigma^-(\mathfrak{W})$ and the rest follows. □

For $k \in \mathbf{Z}$, $\mathfrak{V} \subseteq \mathbf{Z}^n$, we define $k\mathfrak{V} = \{k\underline{v} \mid \underline{v} \in \mathfrak{V}\}$, where $k\underline{v} = (kv_1, ..., kv_n)$. Then it is not hard to prove:

Lemma 2 *If $k \in \mathbf{Z}^+$, $\mathfrak{V} \subseteq \mathbf{Z}^n$, then $\sigma^+(k\mathfrak{V}) = \frac{1}{k^n}\sigma^+(\mathfrak{V})$ and $\sigma^-(k\mathfrak{V}) = \frac{1}{k^n}\sigma^-(\mathfrak{V})$. Also if $\mathfrak{V}$ has a density, so does $k\mathfrak{V}$, and $\sigma(k\mathfrak{V}) = \frac{1}{k^n}\sigma(\mathfrak{V})$.* □

Note also that if $\mathfrak{V}, \mathfrak{W} \subseteq \mathbf{Z}^n$, $\mathfrak{V} \cap \mathfrak{W} = \emptyset$ and $\mathfrak{W}$ has a density, then

$$\begin{aligned}\sigma^+(\mathfrak{V} \cup \mathfrak{W}) &= \overline{\lim_{\rho\to\infty}}\, \frac{|\mathfrak{B}^n_\rho \cap (\mathfrak{V} \cup \mathfrak{W})|}{|\mathfrak{B}^n_\rho|} \\ &= \overline{\lim_{\rho\to\infty}} \left(\frac{|\mathfrak{B}^n_\rho \cap \mathfrak{V}|}{|\mathfrak{B}^n_\rho|} + \frac{|\mathfrak{B}^n_\rho \cap \mathfrak{W}|}{|\mathfrak{B}^n_\rho|} \right) \\ &= \overline{\lim_{\rho\to\infty}}\, \frac{|\mathfrak{B}^n_\rho \cap \mathfrak{V}|}{|\mathfrak{B}^n_\rho|} + \lim_{\rho\to\infty} \frac{|\mathfrak{B}^n_\rho \cap \mathfrak{W}|}{|\mathfrak{B}^n_\rho|} \\ &= \sigma^+(\mathfrak{V}) + \sigma(\mathfrak{W}).\end{aligned}$$

We say $\mathfrak{V} \subseteq \mathbf{Z}^n$ is periodic if its equivalence class under $\sim$, $\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V}\}$, is finite. Clearly if $\mathfrak{V} \subseteq \mathbf{Z}^n$ is periodic and $\mathfrak{W} \sim \mathfrak{V}$ then $\mathfrak{W}$ is periodic. For example $\mathfrak{D}^2$ is periodic as its equivalence class is of order 3. The reason for this terminology is:

Lemma 3 *A subset $\mathfrak{V}$ of $\mathbf{Z}^n$ is periodic if and only if $\exists k \in \mathbf{Z}^+$ with $\mathfrak{V} + k\underline{e}_i = \mathfrak{V}$ for each $i = 1, \ldots, n$, where $\underline{e}_i = (\delta_{i1}, \ldots, \delta_{in})$.*

Proof $\Leftarrow$: If $\mathfrak{V} + k\underline{e}_i = \mathfrak{V}$ for each $i = 1, \ldots, n$, then any $\mathfrak{W} \sim \mathfrak{V}$ is of the form $\mathfrak{V} + \underline{u}$ with $0 \le u_i < k$ for each i, whence $|\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V}\}| \le k^n$.
$\Rightarrow$: For each i, $\{\mathfrak{V} + r\underline{e}_i \mid r \in \mathbf{Z}\} \subseteq \{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V}\}$ which is finite so $\exists r_i, s_i \in \mathbf{Z}$, $r_i < s_i$, with $\mathfrak{V} + r_i\underline{e}_i = \mathfrak{V} + s_i\underline{e}_i$. Then $\mathfrak{V} + (s_i - r_i)\underline{e}_i = \mathfrak{V}$. Now $k = \prod_{i=1}^n (s_i - r_i)$ has the required property. □

Such a k will be called a period of $\mathfrak{V}$. So if $k \in \mathbf{Z}^+$ and $\mathfrak{V}$ is periodic with period ℓ then $k\mathfrak{V}$ has $k\ell$ as a period and so is periodic by Lemma 3. The next result confirms that $\sigma(\mathfrak{D}^2) = \frac{2}{3}$.

Lemma 4 *If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is periodic, then $\mathfrak{V}$ has a density,*

$$\sigma(\mathfrak{V}) = \frac{|\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{0} \in \mathfrak{W}\}|}{|\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V}\}|}.$$

Proof Note that if $\underline{u} \in \mathbf{Z}^n$, $\{\mathfrak{W} - \underline{u} \mid \mathfrak{W} \subseteq \mathbf{Z}^n,\ \mathfrak{W} \sim \mathfrak{V}\} = \{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V}\}$. Thus

$$\begin{aligned} |\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{u} \in \mathfrak{W}\}| &= |\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{0} \in \mathfrak{W} - \underline{u}\}| \\ &= |\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{0} \in \mathfrak{W}\}|. \end{aligned}$$

$$\begin{aligned} \sum_{\mathfrak{W} \sim \mathfrak{V}} \sigma_\rho(\mathfrak{W}) &= \frac{\sum_{\mathfrak{W} \sim \mathfrak{V}} |\mathfrak{B}_\rho^n \cap \mathfrak{W}|}{|\mathfrak{B}_\rho^n|} \\ &= \frac{\sum_{\underline{u} \in \mathfrak{B}_\rho^n} |\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{u} \in \mathfrak{W}\}|}{|\mathfrak{B}_\rho^n|} \\ &= \frac{|\mathfrak{B}_\rho^n| \cdot |\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{0} \in \mathfrak{W}\}|}{|\mathfrak{B}_\rho^n|} \\ &= |\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{0} \in \mathfrak{W}\}|. \end{aligned}$$

However $\sigma_\rho(\mathfrak{W}) = \sigma_\rho(\mathfrak{V}) + O(\rho^{-1})$ by Lemma 1, so

$$\sum_{\mathfrak{W} \sim \mathfrak{V}} \sigma_\rho(\mathfrak{W}) = |\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V}\}| \cdot \sigma_\rho(\mathfrak{V}) + O(\rho^{-1})$$

$$\sigma_\rho(\mathfrak{V}) = \frac{|\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V},\ \underline{0} \in \mathfrak{W}\}|}{|\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{V}\}|} + O(\rho^{-1})$$

and the result follows. □

Note that if $\mathfrak{V}$ and $\mathfrak{W}$ are periodic with periods k, ℓ respectively, then $\mathfrak{V} \cap \mathfrak{W}$ has $k\ell$ as a period and so is periodic by Lemma 3. Also, $\mathbf{Z}^n$ is clearly periodic. Thus the periodic subsets of $\mathbf{Z}^n$ form a base for a topology on $\mathbf{Z}^n$, the open sets being of the form $\bigcup_{\alpha \in A} \mathfrak{V}_\alpha$ with $\mathfrak{V}_\alpha$ periodic $\forall \alpha \in A$. We will call these open sets **recurrent** and the closed sets (i.e. their complements) **corecurrent**. The recurrent subsets are then closed under arbitrary unions and finite intersections. If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is recurrent, say $\mathfrak{V} = \bigcup_{\alpha \in A} \mathfrak{V}_\alpha$, and $\mathfrak{W} \sim \mathfrak{V}$, say $\mathfrak{W} = \mathfrak{V} + \underline{u}$, $\underline{u} \in \mathbf{Z}^n$, then $\mathfrak{W} = \bigcup_{\alpha \in A} (\mathfrak{V}_\alpha + \underline{u})$ and so $\mathfrak{W}$ is recurrent. Thus if $\mathfrak{V} \subseteq \mathbf{Z}^n$ is corecurrent and $\mathfrak{W} \sim \mathfrak{V}$ then $\mathfrak{W}$ is corecurrent.

Note that periodic subsets are both recurrent and corecurrent. The converse is not true though, even for forests: Figure 2 shows the central portion of a forest $\mathfrak{J}^2 \subseteq \mathbf{Z}^2$ which is recurrent and corecurrent but not periodic.

Lemma 5 *A subset $\mathfrak{V}$ of $\mathbf{Z}^n$ is recurrent if and only if $\forall \underline{v} \in \mathfrak{V}$, $\exists k \in \mathbf{Z}^+$ such that $\mathfrak{V} \supseteq k\mathbf{Z}^n + \underline{v}$.*

Proof $\Leftarrow$: We have $\mathfrak{V} = \bigcup_{\underline{v} \in \mathfrak{V}} (k(\underline{v})\mathbf{Z}^n + \underline{v})$ for some $k(\underline{v}) \in \mathbf{Z}^+$, so $\mathfrak{V}$ is recurrent.

$\Rightarrow$: If $\mathfrak{V} = \bigcup_{\alpha \in A} \mathfrak{V}_\alpha$ with $\mathfrak{V}_\alpha$ periodic for all $\alpha \in A$ and $\underline{v} \in \mathfrak{V}$ then $\underline{v} \in \mathfrak{V}_\alpha$ for some α. Let $k \in \mathbf{Z}^+$ be a period of $\mathfrak{V}_\alpha$. Then $\mathfrak{V}_\alpha \supseteq k\mathbf{Z}^n + \underline{v}$, so $\mathfrak{V} \supseteq k\mathbf{Z}^n + \underline{v}$, some $k \in \mathbf{Z}^+$. $\square$

Lemma 6 *If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is non-empty and recurrent, then $\sigma^-(\mathfrak{V}) > 0$.*

Proof By Lemma 5, $\mathfrak{V} \supseteq k\mathbf{Z}^n + \underline{v}$ for some $\underline{v} \in \mathfrak{V}$, $k \in \mathbf{Z}^+$, so $\sigma^-(\mathfrak{V}) \geq \sigma^-(k\mathbf{Z}^n + \underline{v}) = \frac{1}{k^n} > 0$ by Lemmas 1 and 2. $\square$

We now return to the main problem. If $\mathfrak{V} \subseteq \mathbf{Z}^n$ we say $\mathfrak{V}$ is a **jungle** if $\mathbf{Z}^n \setminus \mathfrak{V}$ is totally disconnected (i.e. has no edges) and each component (i.e. maximal set of points in which there is a path between any two) of $\mathfrak{V}$ is infinite. So $\mathfrak{J}^2$ is an example of a jungle. The notion of a jungle is in some sense dual to that of a forest, as is seen from the following theorem, due to Marcus Moore, which half solves the problem.

Theorem 7 *(i) If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a forest, $\sigma^+(\mathfrak{V}) \leq \frac{n}{2n-1}$.*
(ii) If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a jungle, $\sigma^-(\mathfrak{V}) \geq \frac{n}{2n-1}$.
Thus $\mu_n \leq \frac{n}{2n-1}$. Also if $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a forest and a jungle, then $\sigma(\mathfrak{V}) = \frac{n}{2n-1}$.

Proof We will find an expression for the number of edges in $\mathfrak{B}^n_\rho \cap \mathfrak{V}$:

$$\begin{aligned} \mathrm{e}(\mathfrak{B}^n_\rho \cap \mathfrak{V}) &= \frac{1}{2} \sum_{\underline{v} \in \mathfrak{B}^n_\rho \cap \mathfrak{V}} |\{\underline{u} \in \mathfrak{B}^n_\rho \cap \mathfrak{V} \mid \mathrm{d}(\underline{u}, \underline{v}) = 1\}| \\ &= \frac{1}{2} \sum_{\underline{v} \in \mathfrak{B}^n_\rho \cap \mathfrak{V}} |\{\underline{u} \in \mathfrak{V} \mid \mathrm{d}(\underline{u}, \underline{v}) = 1\}| + \mathrm{O}(\rho^{n-1}) \end{aligned}$$

$$= \frac{1}{2}\sum_{i=1}^{n}[|\{\underline{v} \in \mathfrak{B}_\rho^n \cap \mathfrak{V} \mid \underline{v}+\underline{e}_i \in \mathfrak{V}\}| + |\{\underline{v} \in \mathfrak{B}_\rho^n \cap \mathfrak{V} \mid \underline{v}-\underline{e}_i \in \mathfrak{V}\}|]$$

$$+ O(\rho^{n-1})$$

$$= \frac{1}{2}\sum_{i=1}^{n}[|\mathfrak{B}_\rho^n \cap \mathfrak{V} \cap (\mathfrak{V}-\underline{e}_i)| + |\mathfrak{B}_\rho^n \cap \mathfrak{V} \cap (\mathfrak{V}+\underline{e}_i)|] + O(\rho^{n-1})$$

$$= \frac{1}{2}\sum_{i=1}^{n}[|\mathfrak{B}_\rho^n \cap \mathfrak{V}| + |\mathfrak{B}_\rho^n \cap (\mathfrak{V}-\underline{e}_i)| - |\mathfrak{B}_\rho^n| + |\mathfrak{B}_\rho^n \setminus (\mathfrak{V} \cup (\mathfrak{V}-\underline{e}_i))|$$

$$+ |\mathfrak{B}_\rho^n \cap \mathfrak{V}| + |\mathfrak{B}_\rho^n \cap (\mathfrak{V}+\underline{e}_i)| - |\mathfrak{B}_\rho^n| + |\mathfrak{B}_\rho^n \setminus (\mathfrak{V} \cup (\mathfrak{V}+\underline{e}_i))|]$$

$$+ O(\rho^{n-1}).$$

Thus

$$\frac{e(\mathfrak{B}_\rho^n \cap \mathfrak{V})}{|\mathfrak{B}_\rho^n|} = \frac{1}{2}\sum_{i=1}^{n}[\sigma_\rho(\mathfrak{V}-\underline{e}_i) + \sigma_\rho(\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V}-\underline{e}_i))) + \sigma_\rho(\mathfrak{V}+\underline{e}_i)$$

$$+ \sigma_\rho(\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V}+\underline{e}_i)))] + n\sigma_\rho(\mathfrak{V}) - n + O(\rho^{-1})$$

$$= 2n\sigma_\rho(\mathfrak{V}) + \sum_{i=1}^{n}\sigma_\rho(\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V}+\underline{e}_i))) - n + O(\rho^{-1})$$

by Lemma 1.
(i) If $\mathfrak{V}$ is a forest, $\mathfrak{B}_\rho^n \cap \mathfrak{V}$ is a finite forest, so $e(\mathfrak{B}_\rho^n \cap \mathfrak{V}) \le |\mathfrak{B}_\rho^n \cap \mathfrak{V}|$. Thus $\sigma_\rho(\mathfrak{V}) \ge 2n\sigma_\rho(\mathfrak{V}) - n + O(\rho^{-1})$ whence $\sigma_\rho(\mathfrak{V}) \le \frac{n}{2n-1} + O(\rho^{-1})$. So $\sigma^+(\mathfrak{V}) \le \frac{n}{2n-1}$.
(ii) If $\mathfrak{V}$ is a jungle, each component of $\mathfrak{B}_\rho^n \cap \mathfrak{V}$ meets the boundary of $\mathfrak{B}_\rho^n$, so the number of components is $O(\rho^{n-1})$. Hence $e(\mathfrak{B}_\rho^n \cap \mathfrak{V}) \ge |\mathfrak{B}_\rho^n \cap \mathfrak{V}| - O(\rho^{n-1})$. Also $\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V}+\underline{e}_i)) = \emptyset$. Thus $\sigma_\rho(\mathfrak{V}) \le 2n\sigma_\rho(\mathfrak{V}) - n + O(\rho^{-1})$ whence $\sigma_\rho(\mathfrak{V}) \ge \frac{n}{2n-1} + O(\rho^{-1})$. So $\sigma^-(\mathfrak{V}) \ge \frac{n}{2n-1}$.
Hence $\mu_n = \sup\{\sigma^+(\mathfrak{V}) \mid \mathfrak{V} \subseteq \mathbf{Z}^n, \mathfrak{V} \text{ a forest}, \sigma^+(\mathfrak{V}) = \sigma^-(\mathfrak{V})\} \le \frac{n}{2n-1}$. And if $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a forest and a jungle, we have $\frac{n}{2n-1} \le \sigma^-(\mathfrak{V}) \le \sigma^+(\mathfrak{V}) \le \frac{n}{2n-1}$ so $\sigma^-(\mathfrak{V}) = \sigma^+(\mathfrak{V}) = \frac{n}{2n-1}$, i.e. $\sigma(\mathfrak{V}) - \frac{n}{2n-1}$. □

Note, though, that not every forest $\mathfrak{V} \subseteq \mathbf{Z}^n$ with $\sigma^+(\mathfrak{V}) = \frac{n}{2n-1}$ is a jungle, nor is every jungle $\mathfrak{V} \subseteq \mathbf{Z}^n$ with $\sigma^-(\mathfrak{V}) = \frac{n}{2n-1}$ a forest. However:

Lemma 8 *(i) If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a forest with $\sigma^+(\mathfrak{V}) = \frac{n}{2n-1}$ and $\mathfrak{V}$ is corecurrent, then $\mathfrak{V}$ is a jungle.*
(ii) If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a jungle with $\sigma^-(\mathfrak{V}) = \frac{n}{2n-1}$ and $\mathfrak{V}$ is recurrent, then $\mathfrak{V}$ is a forest.
Proof (i) Suppose $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a corecurrent forest with $\sigma^+(\mathfrak{V}) = \frac{n}{2n-1}$ which is not a jungle. Suppose $\mathbf{Z}^n \setminus \mathfrak{V}$ is not totally disconnected. Then $\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V}+\underline{e}_i))$ is non-empty

for some $i = 1, \ldots, n$. But $\mathfrak{V}$ is corecurrent, so $\mathfrak{V} + \underline{e}_i$ is corecurrent, so $\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V} + \underline{e}_i))$ is recurrent. Thus $\sigma^-(\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V} + \underline{e}_i))) > 0$ by Lemma 6. But

$$\sigma_\rho(\mathfrak{V}) \geq 2n\sigma_\rho(\mathfrak{V}) + \sigma_\rho(\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V} + \underline{e}_i))) - n + \mathrm{O}(\rho^{-1}).$$

So $(1-2n)\sigma^+(\mathfrak{V}) + n > 0$, a contradiction. Hence $\mathbf{Z}^n \setminus \mathfrak{V}$ is totally disconnected, so $\mathfrak{V}$ has a finite component, which we will call $\mathfrak{F}$. Let $\mathfrak{N} = \{\underline{v} \in \mathbf{Z}^n \setminus \mathfrak{F} \mid \mathrm{d}(\underline{u}, \underline{v}) = 1 \text{ for some } \underline{u} \in \mathfrak{F}\}$. Then $\mathfrak{N}$ is finite and $\mathfrak{N} \cap \mathfrak{V} = \emptyset$. So $\underline{0} \notin \bigcup_{\underline{v} \in \mathfrak{N}} (\mathfrak{V} - \underline{v})$. But then $\mathbf{Z}^n \setminus \bigcup_{\underline{v} \in \mathfrak{N}} (\mathfrak{V} - \underline{v})$ is recurrent and contains $\underline{0}$, so $\exists k \in \mathbf{Z}^+$ such that $k\mathbf{Z}^n \subseteq \mathbf{Z}^n \setminus \bigcup_{\underline{v} \in \mathfrak{N}} (\mathfrak{V} - \underline{v})$ by Lemma 5, i.e. $\forall \underline{u} \in \mathbf{Z}^n, \forall \underline{v} \in \mathfrak{N}$, $k\underline{u} \notin \mathfrak{V} - \underline{v}$ so $(\mathfrak{N} + k\underline{u}) \cap \mathfrak{V} = \emptyset$. So if we let ℓ be a multiple of k with $\ell \geq |\mathfrak{F}|$, there is a component of $\mathfrak{V}$ contained in $\mathfrak{F} + \ell\underline{u}$ for all $\underline{u} \in \mathbf{Z}^n$ (since $\mathbf{Z}^n \setminus \mathfrak{V}$ is totally disconnected) and the $\mathfrak{F} + \ell\underline{u}$ are disjoint. Then the number of components in $\mathfrak{B}^n_\rho \cap \mathfrak{V} \geq \frac{1}{\ell^n}|\mathfrak{B}^n_\rho| + \mathrm{O}(\rho^{n-1})$ so

$$\mathrm{e}(\mathfrak{B}^n_\rho \cap \mathfrak{V}) \leq |\mathfrak{B}^n_\rho \cap \mathfrak{V}| - \frac{1}{\ell n}|\mathfrak{B}^n_\rho| - \mathrm{O}(\rho^{n-1}).$$

Hence

$$\sigma_\rho(\mathfrak{V}) - \frac{1}{\ell n} \geq 2n\sigma_\rho(\mathfrak{V}) - n + \mathrm{O}(\rho^{-1}),$$

so $(1-2n)\sigma^+(\mathfrak{V}) + n > 0$, another contradiction.

(ii) Suppose $\mathfrak{V} \subseteq \mathbf{Z}^n$ is a recurrent jungle with $\sigma^-(\mathfrak{V}) = \frac{n}{2n-1}$ which is not a forest. Then $\mathfrak{V}$ has a cycle consisting of the points in a finite set $\mathfrak{L}$. Then $\underline{0} \in \bigcap_{\underline{v} \in \mathfrak{L}} (\mathfrak{V} - \underline{v})$ which is recurrent, so $\exists k \in \mathbf{Z}^+$ such that $k\mathbf{Z}^n \subseteq \bigcap_{\underline{v} \in \mathfrak{L}} (\mathfrak{V} - \underline{v})$ by Lemma 5, i.e. $\forall \underline{u} \in \mathbf{Z}^n, \forall \underline{v} \in \mathfrak{L}, k\underline{u} \in \mathfrak{V} - \underline{v}$, so $\mathfrak{L} + k\underline{u} \subseteq \mathfrak{V}$. So if we let ℓ be a multiple of k with $\ell \geq |\mathfrak{L}|$, there are disjoint cycles $\mathfrak{L} + \ell\underline{u} \subseteq \mathfrak{V}$ for all $\underline{u} \in \mathbf{Z}^n$. So

$$\mathrm{e}(\mathfrak{B}^n_\rho \cap \mathfrak{V}) \geq |\mathfrak{B}^n_\rho \cap \mathfrak{V}| + \frac{1}{\ell n}|\mathfrak{B}^n_\rho| + \mathrm{O}(\rho^{n-1}).$$

Also $\mathbf{Z}^n \setminus (\mathfrak{V} \cup (\mathfrak{V} + \underline{e}_i)) = \emptyset$, hence

$$\sigma_\rho(\mathfrak{V}) + \frac{1}{\ell n} \leq 2n\sigma_\rho(\mathfrak{V}) - n + \mathrm{O}(\rho^{-1}),$$

so $(2n-1)\sigma^-(\mathfrak{V}) - n > 0$, a contradiction. □

We can now complete the solution using the following construction, which is due to Mark Owen and the author: let

$$\begin{aligned}\mathfrak{M}^n = \{\underline{v} \in \mathbf{Z}^n \mid\ & v_1 + 2v_2 + \ldots + (n-1)v_{n-1} \equiv 0 \pmod{2n-1} \\ & \text{or } v_1 + \ldots + v_n \equiv 1 \pmod 2\}.\end{aligned}$$

The central portion of $\mathfrak{M}^2$ is shown in Figure 4.

Theorem 9 *For each $n \in \mathbf{Z}^+$, $\mathfrak{M}^n$ is a forest of density $\frac{n}{2n-1}$ (and also a jungle). Thus $\mu_n \geq \frac{n}{2n-1}$.*

Proof First observe $\mathfrak{M}^n$ is periodic, for any translate of $\mathfrak{M}^n$ is of the form

$$\{\underline{v} \in \mathbf{Z}^n \mid v_1 + 2v_2 + \ldots + (n-1)v_{n-1} \equiv a \pmod{2n-1} \\ \text{or } v_1 + \ldots + v_n \equiv b \pmod 2\}.$$

So $|\{\mathfrak{W} \subseteq \mathbf{Z}^n \mid \mathfrak{W} \sim \mathfrak{M}^n\}| \leq 2(2n-1)$. In fact it is easy to see that the $2(2n-1)$ translates $\mathfrak{M}^n + r\underline{e}_1$, $0 \leq r < 2(2n-1)$ are all different, and $\underline{0}$ is contained in $2n-1+1 = 2n$ of them, so $\sigma(\mathfrak{M}^n) = \frac{n}{2n-1}$ by Lemma 4.
Also $\mathfrak{M}^n$ is a jungle. For $\mathbf{Z}^n \setminus \mathfrak{M}^n \subseteq \{\underline{v} \in \mathbf{Z}^n \mid v_1 + \ldots + v_n \equiv 0 \pmod 2\}$ which is totally disconnected, whence so is $\mathbf{Z}^n \setminus \mathfrak{M}^n$. Clearly any $\underline{v} \in \mathfrak{M}^n$ with $v_1 + 2v_2 + \ldots + (n-1)v_{n-1} \equiv 0 \pmod{2n-1}$ is contained in the infinite component $\{(v_1, v_2, \ldots, v_{n-1}, z) \mid z \in \mathbf{Z}\} \subseteq \mathfrak{M}^n$. Otherwise $v_1 + 2v_2 + \ldots + (n-1)v_{n-1} \equiv \pm j \pmod{2n-1}$ for some j, $1 \leq j \leq n-1$. Then $\underline{v}$ is adjacent to $\underline{u}$, where $u_i = v_i \mp \delta_{ij}$. This $\underline{u}$ has $u_1 + 2u_2 + \ldots + (n-1)u_{n-1} = v_1 + 2v_2 + \ldots + (n-1)v_{n-1} \mp j \equiv 0 \pmod{2n-1}$ and so $\underline{v}$ is contained in an infinite component. Hence $\mathfrak{M}^n$ is a forest by Lemma 8 (ii). Thus $\mu_n \geq \frac{n}{2n-1}$. □

So by Theorems 7 and 9, $\mu_n = \frac{n}{2n-1}$. In view of this result, we say a forest $\mathfrak{V} \subseteq \mathbf{Z}^n$ is **maximal** if $\sigma(\mathfrak{V}) = \frac{n}{2n-1}$. We now investigate the maximal forests.

Let $\mathfrak{C}^n = \{\underline{v} \in \mathbf{Z}^n \mid v_1 + \ldots + v_n \equiv 0 \pmod 2\}$, an n-dimensional chequerboard. Then $\mathfrak{C}^n$ has only one other translate, $\mathfrak{C}^n + \underline{e}_1 = \mathbf{Z}^n \setminus \mathfrak{C}^n$, whence $\mathfrak{C}^n$ is periodic and has density $\frac{1}{2}$ by Lemma 4. Note that $2\mathbf{Z}^n \subseteq \mathfrak{C}^n$, so for $\mathfrak{V} \subseteq \mathbf{Z}^n$, define

$$D(\mathfrak{V}) = 2\mathfrak{V} \cup (\mathbf{Z}^n \setminus \mathfrak{C}^n).$$

Then $\sigma^+(D(\mathfrak{V})) = \frac{1}{2^n}\sigma^+(\mathfrak{V}) + \frac{1}{2}$ and $\sigma^-(D(\mathfrak{V})) = \frac{1}{2^n}\sigma^-(\mathfrak{V}) + \frac{1}{2}$ by Lemma 2. The importance of this definition is seen from the following lemma, whose proof is easy:

Lemma 10 *If $\mathfrak{V} \subseteq \mathbf{Z}^n$, there is a bijection between the set of cycles in $\mathfrak{V}$ and the set of cycles in $D(\mathfrak{V})$, with each cycle in $\mathfrak{V}$ corresponding to one twice as long in $D(\mathfrak{V})$.* □

So given a forest $\mathfrak{V} \subseteq \mathbf{Z}^n$, we can form a new forest $D(\mathfrak{V}) \subseteq \mathbf{Z}^n$. We call a subset $\mathfrak{V}$ of $\mathbf{Z}^n$ **primitive** if none of its translates can be obtained in this way, i.e. if $\mathfrak{V} \not\sim D(\mathfrak{U})$ for all $\mathfrak{U} \subseteq \mathbf{Z}^n$. Then clearly if $\mathfrak{V} \subseteq \mathbf{Z}^n$ is primitive and $\mathfrak{W} \sim \mathfrak{V}$, $\mathfrak{W}$ is primitive. If $\mathfrak{V} \subseteq \mathbf{Z}^n$ can be obtained from a primitive by finitely many steps and a translation, we call $\mathfrak{V}$ **elementary**, i.e. if $\mathfrak{V} \sim D^r(\mathfrak{U})$ for some $\mathfrak{U} \subseteq \mathbf{Z}^n$ primitive and some $r \geq 0$. Again, if $\mathfrak{V} \subseteq \mathbf{Z}^n$ is elementary and $\mathfrak{W} \sim \mathfrak{V}$, it is clear that $\mathfrak{W}$ is elementary. We shall be interested in the non-elementary subsets of $\mathbf{Z}^n$.

Lemma 11 *If $\mathfrak{V}, \mathfrak{W} \subseteq \mathbf{Z}^n$, then $D(\mathfrak{V}) \sim D(\mathfrak{W})$ if and only if $\mathfrak{V} \sim \mathfrak{W}$.*

Proof $\Leftarrow$: If $\mathfrak{V} = \mathfrak{W} + \underline{u}$, some $\underline{u} \in \mathbf{Z}^n$, then $D(\mathfrak{V}) = D(\mathfrak{W} + \underline{u}) = 2(\mathfrak{W} + \underline{u}) \cup (\mathbf{Z}^n \setminus \mathfrak{C}^n) = (2\mathfrak{W} + 2\underline{u}) \cup ((\mathbf{Z}^n \setminus \mathfrak{C}^n) + 2\underline{u}) = (2\mathfrak{W} \cup (\mathbf{Z}^n \setminus \mathfrak{C}^n)) + 2\underline{u} = D(\mathfrak{W}) + 2\underline{u}$.
$\Rightarrow$: Suppose $D(\mathfrak{V}) = D(\mathfrak{W}) + \underline{u}$. Then $2\mathfrak{V} \cup (\mathbf{Z}^n \setminus \mathfrak{C}^n) = (2\mathfrak{W} \cup (\mathbf{Z}^n \setminus \mathfrak{C}^n)) + \underline{u} = (2\mathfrak{W} + \underline{u}) \cup ((\mathbf{Z}^n \setminus \mathfrak{C}^n) + \underline{u})$. But $(\mathbf{Z}^n \setminus \mathfrak{C}^n) + \underline{u} = \mathbf{Z}^n \setminus \mathfrak{C}^u$ or $\mathfrak{C}^n$. If $(\mathbf{Z}^n \setminus \mathfrak{C}^n) + \underline{u} = \mathbf{Z}^n \setminus \mathfrak{C}^n$ then

$2\mathfrak{V} = 2\mathfrak{W} + \underline{u}$. If $\underline{u} \neq 2\underline{v}$, some $\underline{v} \in \mathbf{Z}^n$, then $2\mathfrak{V} \cap 2\mathbf{Z}^n = 2\mathfrak{W} \cap 2\mathbf{Z}^n = \emptyset$ so $\mathfrak{V} = \mathfrak{W} = \emptyset$. If $\underline{u} = 2\underline{v}$, some $\underline{v} \in \mathbf{Z}^n$, then $\mathfrak{V} = \mathfrak{W} + \underline{v}$ so $\mathfrak{V} \sim \mathfrak{W}$. Otherwise $(\mathbf{Z}^n \setminus \mathfrak{C}^n) + \underline{u} = \mathfrak{C}^n$, so $2\mathfrak{V} = \mathfrak{C}^n$ and $2\mathfrak{W} + \underline{u} = \mathbf{Z}^n \setminus \mathfrak{C}^n$, so $2\mathfrak{W} = (\mathbf{Z}^n \setminus \mathfrak{C}^n) - \underline{u} = \mathfrak{C}^n$, whence $\mathfrak{V} = \mathfrak{W}$. □

Lemma 12 *A subset $\mathfrak{V}$ of $\mathbf{Z}^n$ is non-elementary if and only if for each $r \geq 0$, $\exists \mathfrak{U} \subseteq \mathbf{Z}^n$ with $\mathfrak{V} \sim D^r(\mathfrak{U})$.*

Proof $\Leftarrow$: Suppose $\mathfrak{V}$ is elementary. Then $\mathfrak{V} \sim D^r(\mathfrak{W})$ for some $\mathfrak{W} \subseteq \mathbf{Z}^n$ primitive. If $\mathfrak{V} \sim D^{r+1}(\mathfrak{U})$, then $D^r(\mathfrak{W}) \sim D^r(D(\mathfrak{U}))$ so $\mathfrak{W} \sim D(\mathfrak{U})$ by Lemma 11 applied r times, contradicting $\mathfrak{W}$ primitive.

$\Rightarrow$: Suppose $\mathfrak{V} \not\sim D^r(\mathfrak{U})$ for all $\mathfrak{U} \subseteq \mathbf{Z}^n$, some $r \geq 0$. Take r to be the least such. Clearly $r \geq 1$, as $\mathfrak{V} \sim D^0(\mathfrak{V})$, so $\mathfrak{V} \sim D^{r-1}(\mathfrak{W})$, some $\mathfrak{W} \subseteq \mathbf{Z}^n$. But now $\mathfrak{W}$ is primitive, for if $\mathfrak{W} \sim D(\mathfrak{U})$ for some $\mathfrak{U} \subseteq \mathbf{Z}^n$, we would have $\mathfrak{V} \sim D^r(\mathfrak{U})$, a contradiction. □

The next lemma shows why we are interested in these subsets.

Lemma 13 *If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is non-elementary then $\mathfrak{V}$ is a forest.*

Proof Suppose $\mathfrak{V} \subseteq \mathbf{Z}^n$ is non-elementary and contains a cycle of length m. Take r such that $2^r > m$. By Lemma 12, $\mathfrak{V} \sim D^r(\mathfrak{U})$ for some $\mathfrak{U} \subseteq \mathbf{Z}^n$. But then $\mathfrak{U}$ contains a cycle of length $\frac{m}{2^r}$ by Lemma 10, which is absurd. □

If $\mathfrak{V} \subseteq \mathbf{Z}^n$ has $D(\mathfrak{V}) \sim \mathfrak{V}$ we say that $\mathfrak{V}$ is pretty. If so, then by induction, $D^r(\mathfrak{V}) \sim \mathfrak{V}$, for all $r \geq 0$, so $\mathfrak{V}$ is non-elementary by Lemma 12. If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is pretty and $\mathfrak{W} \sim \mathfrak{V}$ then $D(\mathfrak{W}) \sim D(\mathfrak{V}) \sim \mathfrak{V} \sim \mathfrak{W}$, so $\mathfrak{W}$ is pretty. Subsets of $\mathbf{Z}^n$ which are non-elementary but not pretty are called **grotesque**. The reader can verify that the subset $\mathfrak{G}^2$ of $\mathbf{Z}^2$ whose central portion is shown in Figure 3 is an example of a grotesque subset.

For each $n \in \mathbf{Z}^+$, define $\mathfrak{P}^n = \bigcup_{r\geq 0} D^r(\emptyset)$ and $\mathfrak{Q}^n = \bigcap_{r\geq 0} D^r(\mathbf{Z}^n)$. Note that if $\mathfrak{V}$ is periodic, then $D(\mathfrak{V})$ is, so $D^r(\mathfrak{V})$ is periodic for $r \geq 0$ by induction. Thus $\mathfrak{P}^n$ is recurrent and $\mathfrak{Q}^n$ is corecurrent. The central portion of $\mathfrak{P}^2$ is shown in Figure 6. We have $D(\mathfrak{P}^n) = D(\bigcup_{r\geq 0} D^r(\emptyset)) = \bigcup_{r\geq 0} D^{r+1}(\emptyset) = \mathfrak{P}^n$, and $D(\mathfrak{Q}^n) = D(\bigcap_{r\geq 0} D^r(\mathbf{Z}^n)) = \bigcap_{r\geq 0} D^{r+1}(\mathbf{Z}^n) = \mathfrak{Q}^n$. So both $\mathfrak{P}^n$ and $\mathfrak{Q}^n$ are pretty. Conversely:

Lemma 14 *If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is pretty then $\mathfrak{V} \sim \mathfrak{P}^n$ or $\mathfrak{V} \sim \mathfrak{Q}^n$. If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is grotesque then $\mathfrak{V}$ is recurrent and corecurrent.*

Proof Suppose $\mathfrak{V} \subseteq \mathbf{Z}^n$ is non-elementary. Then by Lemma 12, $\exists \mathfrak{U}_0, \mathfrak{U}_1, \ldots \subseteq \mathbf{Z}^n$ with $\mathfrak{V} \sim D^r(\mathfrak{U}_r)$ for each $r \geq 0$. Since $D^r(\mathfrak{U}_r) \sim \mathfrak{V} \sim D^{r+1}(\mathfrak{U}_{r+1})$, we have $\mathfrak{U}_r \sim D(\mathfrak{U}_{r+1})$ by Lemma 11 applied r times; say $\mathfrak{U}_r = D(\mathfrak{U}_{r+1}) + \underline{v}_r$. Now for each $r \geq 0$, let

$$\mathfrak{X}_r = D(D(\ldots(D(\emptyset) + \underline{v}_{r-1})\ldots) + \underline{v}_1) + \underline{v}_0,$$

understanding $\mathfrak{X}_0 = \emptyset$, and let

$$\mathfrak{Y}_r = D(D(\ldots(D(\mathbf{Z}^n) + \underline{v}_{r-1})\ldots) + \underline{v}_1) + \underline{v}_0,$$

understanding $\mathfrak{Y}_0 = \mathbf{Z}^n$. Then $\forall r \geq 0, \mathfrak{X}_r \subseteq \mathfrak{V} \subseteq \mathfrak{Y}_r$, so $\bigcup_{r\geq 0} \mathfrak{X}_r \subseteq \mathfrak{V} \subseteq \bigcap_{r\geq 0} \mathfrak{Y}_r$. Since $\mathfrak{X}_r$ and $\mathfrak{Y}_r$ are periodic for all $r \geq 0$, $\bigcup_{r\geq 0} \mathfrak{X}_r$ is recurrent and $\bigcap_{r\geq 0} \mathfrak{X}_r$ is corecurrent. Now suppose

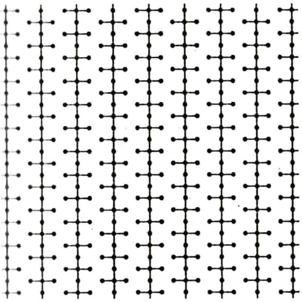


Figure 4 $\mathfrak{M}^2$

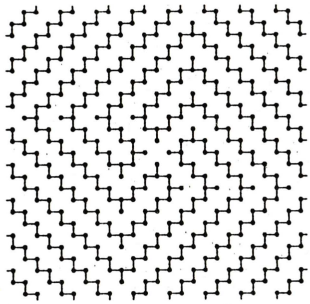


Figure 5 $\mathfrak{S}^2$

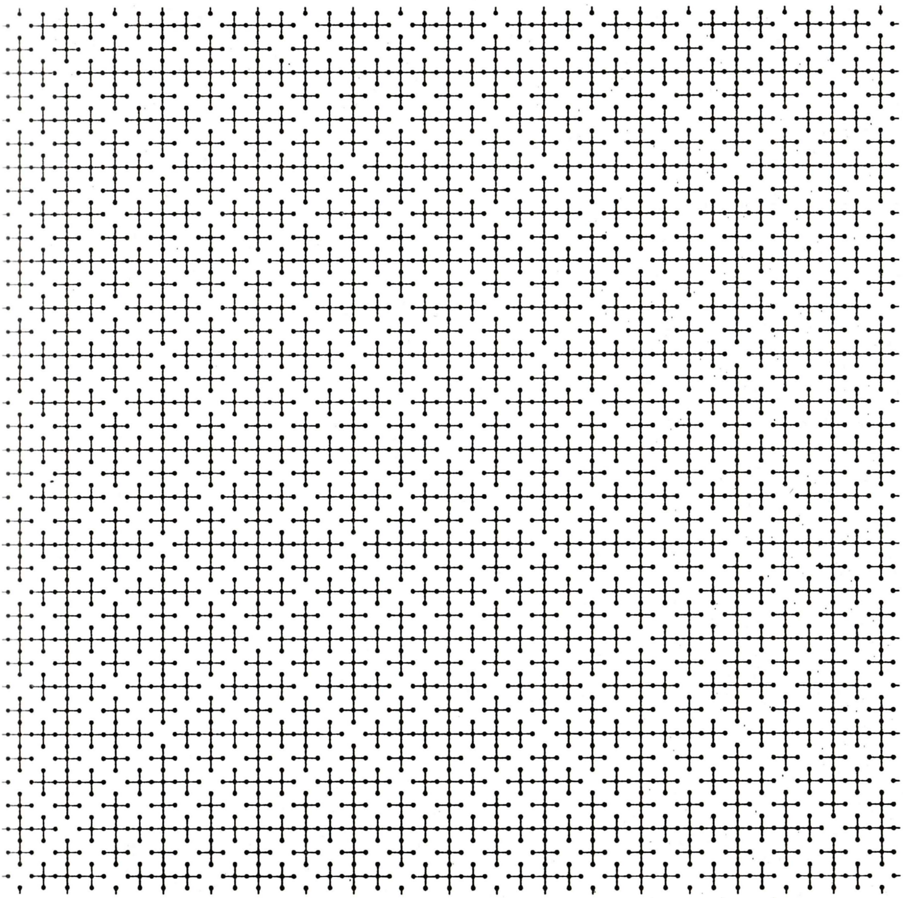


Figure 6 $\mathfrak{D}^2$

$\underline{u} \in \bigcap_{r\geq 0} \mathfrak{Y}_r \setminus \bigcup_{r\geq 0} \mathfrak{X}_r$. Then $\underline{u} \in \mathfrak{Y}_r$ and $\underline{u} \notin \mathfrak{X}_r$ for all $r \geq 0$, i.e. $\underline{u} \in D^r(\mathbf{Z}^n)+2^{r-1}\underline{v}_{r-1}+\ldots+2\underline{v}_1+\underline{v}_0$ and $\underline{u} \notin D^r(\emptyset)+2^{r-1}\underline{v}_{r-1}+\ldots+2\underline{v}_1+\underline{v}_0$, so that $\underline{u} \in 2^r\mathbf{Z}^n+2^{r-1}\underline{v}_{r-1}+\ldots+2\underline{v}_1+\underline{v}_0$ for all $r \geq 0$. So there can be at most one such $\underline{u}$. If one exists, then $\exists \underline{u}_0, \underline{u}_1, \ldots \in \mathbf{Z}^n$ with $\underline{u}-2^{r-1}\underline{v}_{r-1}-\ldots-2\underline{v}_1-\underline{v}_0 = 2^r\underline{u}_r$. So $\mathfrak{U}_r - \underline{u}_r = D(\mathfrak{U}_{r+1})+\underline{v}_r-\underline{u}_r = D(\mathfrak{U}_{r+1})-2\underline{u}_{r+1} = D(\mathfrak{U}_{r+1}-\underline{u}_{r+1})$. So by the same argument $\bigcup_{r\geq 0} D^r(\emptyset) \subseteq \mathfrak{U}_0 - \underline{u}_0 \subseteq \bigcap_{r\geq 0} D^r(\mathbf{Z}^n)$ with the union and the intersection differing in at most one place, so $\mathfrak{V} - \underline{u} = \mathfrak{U}_0 - \underline{u}_0 = \mathfrak{P}^n$ or $\mathfrak{Q}^n$, so $\mathfrak{V}$ is pretty. Thus if $\mathfrak{V}$ is grotesque, $\mathfrak{V} = \bigcup_{r\geq 0} \mathfrak{X}_r = \bigcap_{r\geq 0} \mathfrak{Y}_r$ and so $\mathfrak{V}$ is recurrent and corecurrent. If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is pretty, $D(\mathfrak{V}) = \mathfrak{V} + \underline{v}$, some $\underline{v} \in \mathbf{Z}^n$, so $\mathfrak{V} - \underline{v} = D(\mathfrak{V}) - 2\underline{v} = D(\mathfrak{V} - \underline{v})$ so as before $\mathfrak{V} - \underline{v} = \mathfrak{P}^n$ or $\mathfrak{Q}^n$. □

Note that we cannot have $\mathfrak{P}^n$ corecurrent and $\mathfrak{Q}^n$ recurrent, for $\underline{0} \in \mathfrak{Q}^n$ but $\underline{0} \notin \mathfrak{P}^n$ so by the above argument this is the only difference, i.e. $\mathfrak{Q}^n \setminus \mathfrak{P}^n = \{\underline{0}\}$ which would have to be recurrent, contradicting Lemma 5. Thus $\mathfrak{Q}^n \not\sim \mathfrak{P}^n$, so there are exactly two classes of pretty subsets of each $\mathbf{Z}^n$.

Lemma 15 *If $\mathfrak{V} \subseteq \mathbf{Z}^n$ is non-elementary then $\sigma(\mathfrak{V}) = \frac{2^{n-1}}{2^n-1}$.*
Proof Observe that if $\mathfrak{W} \subseteq \mathbf{Z}^n$,

$$\begin{aligned}\sigma^+(D(\mathfrak{W})) - \frac{2^{n-1}}{2^n-1} &= \frac{1}{2^n}\sigma^+(\mathfrak{W}) + \frac{1}{2} - \frac{2^{n-1}}{2^n-1}\\ &= \frac{1}{2^n}\left[\sigma^+(\mathfrak{W}) + \frac{2^{n-1}(2^n-1) - 2^n \cdot 2^{n-1}}{2^n-1}\right]\\ &= \frac{1}{2^n}\left[\sigma^+(\mathfrak{W}) - \frac{2^{n-1}}{2^n-1}\right].\end{aligned}$$

Suppose $\sigma^+(\mathfrak{V}) = \frac{2^{n-1}}{2^n-1} + \epsilon$, some $\epsilon > 0$. By Lemma 12, for each $r \geq 0$, $\exists \mathfrak{U} \subseteq \mathbf{Z}^n$ such that $\mathfrak{V} \sim D^r(\mathfrak{U})$. Then $\epsilon = \frac{1}{2^{rn}}[\sigma^+(\mathfrak{U}) - \frac{2^{n-1}}{2^n-1}]$, so $\sigma^+(\mathfrak{U}) = \frac{2^{n-1}}{2^n-1} + 2^{rn}\epsilon$, which is greater than 1 for sufficiently large r, a contradiction. Thus $\sigma^+(\mathfrak{V}) \leq \frac{2^{n-1}}{2^n-1}$. Similarly, $\sigma^-(\mathfrak{V}) \geq \frac{2^{n-1}}{2^n-1}$, so $\sigma^-(\mathfrak{V}) = \sigma^+(\mathfrak{V}) = \frac{2^{n-1}}{2^n-1}$, i.e. $\sigma(\mathfrak{V}) = \frac{2^{n-1}}{2^n-1}$. □

So these non-elementary forests are maximal when $\frac{2^{n-1}}{2^n-1} = \frac{n}{2n-1}$, i.e. when $2^{n-1}(2n-1) = n(2^n-1)$, i.e. when $n = 2^{n-1}$, i.e. for $n = 1$ or 2 only. We will now consider these two cases individually.

In the case $n = 1$, all subsets $\mathfrak{V}$ of $\mathbf{Z}$ are forests, so maximal forests have density 1, as predicted. (In fact, $\mathfrak{M}^1 = \mathbf{Z}$.) For $\mathfrak{V} \subseteq \mathbf{Z}$ to be a jungle, it is necessary and sufficient that $\mathbf{Z} \setminus \mathfrak{V}$ has at most one element, so there are two classes of jungles, the singleton class consisting of $\mathbf{Z}$ and the class of translates of $\mathbf{Z} \setminus \{\underline{0}\}$. Since $\mathfrak{Q}^1 = \bigcap_{r\geq 0} D^r(\mathbf{Z}) = \bigcap_{r\geq 0} \mathbf{Z} = \mathbf{Z}$, we have $\mathfrak{P}^1 = \mathbf{Z} \setminus \{\underline{0}\}$, so the pretty subsets of $\mathbf{Z}$ are precisely the jungles. For $\mathfrak{V} \subseteq \mathbf{Z}$ to be corecurrent and have density 1, we need $\mathfrak{V} = \mathbf{Z}$ by Lemma 6, and so $\mathfrak{V}$ is pretty. Hence by Lemma 14, no subsets of $\mathbf{Z}$ are grotesque.

This enables us to prove the general result:

Lemma 16 *If $\mathfrak{V} \subseteq Z^n$ is grotesque then $\mathfrak{V}$ is not periodic.*
Proof Suppose $\mathfrak{V} \subseteq Z^n$ is grotesque and periodic. By Lemma 3, $\exists k \in Z^+$ such that $\mathfrak{V} + k\underline{e}_1 = \mathfrak{V}$. Now let $k = 2^r\ell$ with $r \geq 0$, ℓ odd. By Lemma 12, $\mathfrak{V} \sim D^r(\mathfrak{U})$ for some $\mathfrak{U} \subseteq Z^n$, say $\mathfrak{V} = D^r(\mathfrak{U}) + \underline{v}$. This $\mathfrak{U}$ is not elementary, for if $\mathfrak{U} \sim D^s(\mathfrak{W})$ with $\mathfrak{W}$ primitive, then $\mathfrak{V} \sim D^{r+s}(\mathfrak{W})$, so $\mathfrak{V}$ is elementary, a contradiction. Also $\mathfrak{U}$ is not pretty, for if $\mathfrak{U} \sim D(\mathfrak{U})$, then $\mathfrak{V} \sim D^r(\mathfrak{U}) \sim D^r(D(\mathfrak{U})) = D(D^r(\mathfrak{U})) \sim D(\mathfrak{V})$, so $\mathfrak{V}$ is pretty, a contradiction. Thus $\mathfrak{U}$ is grotesque. By the above remarks $n > 1$ and so $\sigma(\mathfrak{U}) < 1$ by Lemma 15. However $D^r(\mathfrak{U} + \ell\underline{e}_1) = D^r(\mathfrak{U}) + 2^r\ell\underline{e}_1 = \mathfrak{V} - \underline{v} + k\underline{e}_1 = \mathfrak{V} - \underline{v} = D^r(\mathfrak{U})$, so $\mathfrak{U} + \ell\underline{e}_1 = \mathfrak{U}$. But $\mathfrak{C}^n \pm \ell\underline{e}_1 = Z^n \setminus \mathfrak{C}^n$ as ℓ is odd, so $\mathfrak{U} \supseteq \mathfrak{C}^n$ and $\mathfrak{U} \supseteq Z^n \setminus \mathfrak{C}^n$, so $\mathfrak{U} = Z^n$. Thus $\sigma(\mathfrak{U}) = 1$, a contradiction. $\square$

In the case $n = 2$, for $\mathfrak{V} \subseteq Z^2$, let $F(\mathfrak{V}) = Z^2 \setminus \{(v_1 + v_2, v_1 - v_2) \mid \underline{v} \in \mathfrak{V}\}$. Then

$$\begin{aligned} F^2(\mathfrak{V}) &= Z^2 \setminus \{(u_1 + u_2, u_1 - u_2) \mid \underline{u} \in Z^2 \setminus \{(v_1 + v_2, v_1 - v_2) \mid \underline{v} \in \mathfrak{V}\}\} \\ &= Z^2 \setminus \{(u_1 + u_2, u_1 - u_2) \mid \underline{u} \in Z^2\} \\ &\quad \cup \{((v_1 + v_2) + (v_1 - v_2), (v_1 + v_2) - (v_1 - v_2)) \mid \underline{v} \in \mathfrak{V}\} \\ &= (Z^2 \setminus \mathfrak{C}^2) \cup 2\mathfrak{V} \\ &= D(\mathfrak{V}). \end{aligned}$$

Note that if $\mathfrak{V} \subseteq Z^2, \underline{v} \in Z^2$, $F(\mathfrak{V} + \underline{v}) = Z^2 \setminus \{(u_1 + u_2, u_1 - u_2) \mid \underline{u} \in \mathfrak{V} + \underline{v}\} = Z^2 \setminus \{((u_1 + v_1) + (u_2 + v_2), (u_1 + v_1) - (u_2 + v_2)) \mid \underline{u} \in \mathfrak{V}\} = (Z^2 \setminus \{(u_1 + u_2, u_1 - u_2) \mid \underline{u} \in \mathfrak{V}\}) + (v_1 + v_2, v_1 - v_2) = F(\mathfrak{V}) + (v_1 + v_2, v_1 - v_2)$. So if $\mathfrak{V} \sim \mathfrak{W}$ then $F(\mathfrak{V}) \sim F(\mathfrak{W})$. Also, if $\mathfrak{V}$ is periodic, so is $F(\mathfrak{V})$. Thus if $\mathfrak{V}$ is recurrent, say $\mathfrak{V} = \bigcup_{\alpha \in A} \mathfrak{V}_\alpha$, then $F(\mathfrak{V}) = \bigcap_{\alpha \in A} F(\mathfrak{V}_\alpha)$, so $F(\mathfrak{V})$ is corecurrent. Similarly if $\mathfrak{V}$ is corecurrent, $F(\mathfrak{V})$ is recurrent.

Note that if $\mathfrak{V} \subseteq Z^2$, $\mathfrak{V} = F(\mathfrak{U})$ for some $\mathfrak{U} \subseteq Z^2$ if and only if $Z^2 \setminus \mathfrak{C}^2 \subseteq \mathfrak{V}$. We can now give a more useful characterisation of elementary subsets of Z^2.

Lemma 17 *A subset $\mathfrak{V}$ of Z^2 is elementary if and only if $\mathfrak{V} \sim F^s(\mathfrak{U})$ for some $s \geq 0$ and $\mathfrak{U} \subseteq Z^2$ with $\mathfrak{C}^2 \not\subseteq \mathfrak{U}$ and $Z^2 \setminus \mathfrak{C}^2 \not\subseteq \mathfrak{U}$.*
Proof The condition $\mathfrak{C}^2 \not\subseteq \mathfrak{U}$ and $Z^2 \setminus \mathfrak{C}^2 \not\subseteq \mathfrak{U}$ is clearly equivalent to the condition $\mathfrak{U} \not\sim F(\mathfrak{W})$ for any $\mathfrak{W} \subseteq Z^2$. So:
$\Leftarrow$: Suppose $\mathfrak{V} \sim F^s(\mathfrak{U})$ with $\mathfrak{U}$ satisfying the condition. Then $\mathfrak{U}$ is primitive, for if $\mathfrak{U} \sim D(\mathfrak{W})$ then $\mathfrak{U} \sim F(F(\mathfrak{W}))$, a contradiction. Also, $F(\mathfrak{U})$ is primitive, for if $F(\mathfrak{U}) \sim D(\mathfrak{W})$ then $F(F(\mathfrak{U})) \sim F(D(\mathfrak{W}))$, so $D(\mathfrak{U}) \sim D(F(\mathfrak{W}))$. Then by Lemma 11, $\mathfrak{U} \sim F(\mathfrak{W})$, a contradiction. But $s = 2r$ or $s = 2r + 1$ for some $r \geq 0$, whence $\mathfrak{V} \sim D^r(\mathfrak{U})$ or $\mathfrak{V} \sim D^r(F(\mathfrak{U}))$ so $\mathfrak{V}$ is elementary.
$\Rightarrow$: Suppose $\mathfrak{V} \subseteq Z^2$ is elementary. Then $\exists \mathfrak{U} \subseteq Z^2$, primitive, and $r \geq 0$ such that $\mathfrak{V} \sim D^r(\mathfrak{U}) = F^{2r}(\mathfrak{U})$. If $\mathfrak{U} \not\sim F(\mathfrak{W})$ we are done, so we may assume $\mathfrak{U} \sim F(\mathfrak{W})$. Now $\mathfrak{V} \sim F^{2r+1}(\mathfrak{W})$ and $\mathfrak{W} \not\sim F(\mathfrak{X})$, for if $\mathfrak{W} \sim F(\mathfrak{X})$ we would have $\mathfrak{U} \sim F(\mathfrak{W}) \sim D(\mathfrak{X})$, contradicting $\mathfrak{U}$ primitive. $\square$

Observe for example that $\mathfrak{M}^2 = F(\mathfrak{D}^2)$, with $\mathfrak{C}^2 \not\subseteq \mathfrak{D}^2$ and $Z^2 \setminus \mathfrak{C}^2 \not\subseteq \mathfrak{D}^2$.

Lemma 18 *If $\mathfrak{V} \subseteq Z^2$, $\sigma^+(F(\mathfrak{V})) = 1 - \frac{1}{2}\sigma^-(\mathfrak{V})$ and $\sigma^-(F(\mathfrak{V})) = 1 - \frac{1}{2}\sigma^+(\mathfrak{V})$.*

Proof Observe that

$$\begin{aligned}
|\mathfrak{B}^2_\rho \cap F(\mathfrak{V})| &= |\mathfrak{B}^2_\rho| - |\mathfrak{B}^2_\rho \cap \{(v_1+v_2, v_1-v_2) \mid \underline{v} \in \mathfrak{V}\}| \\
&= |\mathfrak{B}^2_\rho| - |\{\underline{v} \in \mathfrak{V} \mid \sqrt{(v_1+v_2)^2 + (v_1-v_2)^2} < \rho\}| \\
&= |\mathfrak{B}^2_\rho| - |\{\underline{v} \in \mathfrak{V} \mid \sqrt{2(v_1^2+v_2^2)} < \rho\}| \\
&= |\mathfrak{B}^2_\rho| - |\mathfrak{B}^2_{\frac{\rho}{\sqrt{2}}} \cap \mathfrak{V}|.
\end{aligned}$$

$$\begin{aligned}
\sigma^+(F(\mathfrak{V})) = \overline{\lim_{\rho\to\infty}} \frac{|\mathfrak{B}^2_\rho \cap F(\mathfrak{V})|}{|\mathfrak{B}^2_\rho|} &= \overline{\lim_{\rho\to\infty}} \left(\frac{|\mathfrak{B}^2_\rho|}{|\mathfrak{B}^2_\rho|} - \frac{|\mathfrak{B}^2_{\frac{\rho}{\sqrt{2}}} \cap \mathfrak{V}| \cdot |\mathfrak{B}^2_{\frac{\rho}{\sqrt{2}}}|}{|\mathfrak{B}^2_{\frac{\rho}{\sqrt{2}}}| \cdot |\mathfrak{B}^2_\rho|} \right) \\
&= 1 - \left(\underline{\lim_{\rho\to\infty}} \frac{|\mathfrak{B}^2_{\frac{\rho}{\sqrt{2}}} \cap \mathfrak{V}|}{|\mathfrak{B}^2_{\frac{\rho}{\sqrt{2}}}|} \right) \cdot \left(\lim_{\rho\to\infty} \frac{|\mathfrak{B}^2_{\frac{\rho}{\sqrt{2}}}|}{|\mathfrak{B}^2_\rho|} \right) \\
&= 1 - \frac{1}{2}\sigma^-(\mathfrak{V}).
\end{aligned}$$

The rest of the result is proved similarly. □

This gives $\sigma^+(D(\mathfrak{V})) = 1 - \frac{1}{2}\sigma^-(F(\mathfrak{V})) = 1 - \frac{1}{2}(1 - \frac{1}{2}\sigma^+(\mathfrak{V})) = \frac{1}{4}\sigma^+(\mathfrak{V}) + \frac{1}{2}$ and similarly $\sigma^-(D(\mathfrak{V})) = \frac{1}{4}\sigma^-(\mathfrak{V}) + \frac{1}{2}$, as expected.

If $\mathfrak{V} \subseteq \mathbf{Z}^2$ is pretty, $D(\mathfrak{V}) \sim \mathfrak{V}$, so $D(F(\mathfrak{V})) = F(D(\mathfrak{V})) \sim F(\mathfrak{V})$, so $F(\mathfrak{V})$ is pretty. In fact, $F(\mathfrak{P}^2) = F(\bigcup_{r\geq 0} D^r(\emptyset)) = \bigcap_{r\geq 0} F(D^r(\emptyset)) = \bigcap_{r\geq 0} D^r(F(\emptyset)) = \bigcap_{r\geq 0} D^r(\mathbf{Z}^2) = \mathfrak{Q}^2$. And $F(\mathfrak{Q}^2) = F(F(\mathfrak{P}^2)) = D(\mathfrak{P}^2) = \mathfrak{P}^2$. Hence $\mathfrak{P}^2$ is not corecurrent, nor is $\mathfrak{Q}^2$ recurrent, for if either held, then they would both hold, which was shown above to be impossible.

Lemma 19 *(i) A subset $\mathfrak{V}$ of $\mathbf{Z}^2$ is a jungle if and only if $F(\mathfrak{V})$ is a forest.*
(ii) A subset $\mathfrak{V}$ of $\mathbf{Z}^2$ is a forest if and only if $F(\mathfrak{V})$ is a jungle.

Proof (i)$\Leftarrow$: Suppose $\mathfrak{V} \subseteq \mathbf{Z}^2$ is not a jungle. Then either $\mathbf{Z}^2 \setminus \mathfrak{V}$ is totally disconnected, or $\mathfrak{V}$ has a finite component. In the first case, $\exists \underline{v} \in \mathbf{Z}^2 \setminus \mathfrak{V}$ such that $\underline{v} + \underline{e}_1 \in \mathbf{Z}^2 \setminus \mathfrak{V}$ or $\underline{v} + \underline{e}_2 \in \mathbf{Z}^2 \setminus \mathfrak{V}$. Then $(v_1+v_2, v_1-v_2) \in F(\mathfrak{V})$ and either $(v_1+v_2+1, v_1-v_2+1) \in F(\mathfrak{V})$ or $(v_1+v_2+1, v_1-v_2-1) \in F(\mathfrak{V})$. So $F(\mathfrak{V})$ contains either the cycle (v_1+v_2, v_1-v_2), (v_1+v_2+1, v_1-v_2), (v_1+v_2+1, v_1-v_2-1), (v_1+v_2, v_1-v_2-1) or the cycle (v_1+v_2, v_1-v_2), (v_1+v_2+1, v_1-v_2), (v_1+v_2+1, v_1-v_2+1), (v_1+v_2, v_1-v_2+1). Hence $F(\mathfrak{V})$ is not a forest. In the second case, let $\mathfrak{F}$ be a finite component of $\mathfrak{V}$. Then $\{\underline{v} \in F(\mathfrak{V}) \mid d((u_1+u_2, u_1-u_2), \underline{v}) = 1 \text{ for some } \underline{u} \in \mathfrak{F}\}$ contains a cycle, so $F(\mathfrak{V})$ is not a forest.

$\Rightarrow$: Suppose $\mathfrak{V} \subseteq \mathbf{Z}^2$ with $F(\mathfrak{V})$ not a forest. Let $\mathfrak{L}$ be the set of points of a smallest cycle of $F(\mathfrak{V})$. Then $|\mathfrak{L}| \geq 4$. If $|\mathfrak{L}| = 4$, then $\mathfrak{L} = \{(v_1, v_2), (v_1, v_2+1), (v_1+1, v_2+1), (v_1+1, v_2)\}$ for some $\underline{v} \in \mathbf{Z}^2$. If $v_1 \equiv v_2 \pmod 2$ then $v_1+v_2 = 2u_1$, $v_1 - v_2 = 2u_2$ for some $\underline{u} \in \mathbf{Z}^2$. Then $\{(u_1+u_2, u_1-u_2), (u_1+u_2+1, u_1-u_2+1)\} \subseteq \mathfrak{L} \subseteq F(\mathfrak{V})$, so $\{(u_1, u_2), (u_1+1, u_2)\} \subseteq \mathbf{Z}^2 \setminus \mathfrak{V}$. Similarly if $v_1 \not\equiv v_2 \pmod 2$ then $v_1+v_2+1 = 2u_1$, $v_1-v_2-1 = 2u_2$ for some $\underline{u} \in \mathbf{Z}^2$. Then $\{(u_1+u_2, u_1-u_2), (u_1+u_2+1, u_1-u_2-1)\} \subseteq \mathfrak{L} \subseteq F(\mathfrak{V})$, so $\{(u_1, u_2), (u_1, u_2+1)\} \subseteq \mathbf{Z}^2 \setminus \mathfrak{V}$. Thus $\mathbf{Z}^2 \setminus \mathfrak{V}$ is not totally disconnected. If $|\mathfrak{L}| > 4$, let $\mathfrak{H}$ be the set of points of $\mathbf{Z}^2 \setminus F(\mathfrak{V})$

contained inside $\mathfrak{L}$. Then $\mathfrak{H} = \{(v_1 + v_2, v_1 - v_2) \mid \underline{v} \in \mathfrak{F}\}$ for some finite $\mathfrak{F} \subseteq \mathfrak{V}$ and it is not hard to see that $\mathfrak{F}$ is a component of $\mathfrak{V}$. Hence $\mathfrak{V}$ is not a jungle.
(ii) By Lemma 10, $\mathfrak{V} \subseteq \mathbf{Z}^2$ is a forest if and only if $D(\mathfrak{V}) = F(F(\mathfrak{V}))$ is a forest, so by (i), if and only if $F(\mathfrak{V})$ is a jungle. □

This result can be used with Lemma 18 to show the equivalence of parts (i) and (ii) of Theorem 7 and of Lemma 8 in the case $n = 2$. Also, it shows that any non-elementary subset $\mathfrak{V}$ of $\mathbf{Z}^2$ is a jungle, for by Lemma 17, $\mathfrak{V} \sim F(\mathfrak{U})$ with $\mathfrak{U} \subseteq \mathbf{Z}^2$ non-elementary, so a forest by Lemma 13. Then $\sigma(\mathfrak{V}) = \frac{2}{3}$ by Theorem 7, verifying Lemma 15 in this case.

If $\mathfrak{V} \subseteq \mathbf{Z}^2$, let $R(\mathfrak{V}) = \{(-v_2, v_1) \mid \underline{v} \in \mathfrak{V}\}$, i.e. the image of $\mathfrak{V}$ under an anticlockwise rotation of 90° about the origin, and say $\mathfrak{V}$ is **rotational** if $\mathfrak{V} \sim R(\mathfrak{V})$. Then if $\mathfrak{V} \subseteq \mathbf{Z}^2$ is rotational and $\mathfrak{W} \sim \mathfrak{V}$, say $\mathfrak{W} = \mathfrak{V} + \underline{u}$, we have $R(\mathfrak{W}) = \{(-v_2, v_1) \mid \underline{v} \in \mathfrak{V} + \underline{u}\} = \{(-(v_2 + u_2), (v_1 + u_1)) \mid \underline{v} \in \mathfrak{V}\} = \{(-v_2, v_1) \mid \underline{v} \in \mathfrak{V}\} + (-u_2, u_1) = R(\mathfrak{V}) + (-u_2, u_1) \sim R(\mathfrak{V}) \sim \mathfrak{V} \sim \mathfrak{W}$, so $\mathfrak{W}$ is rotational. Clearly R preserves the properties of being periodic, recurrent and corecurrent. Figure 5 shows the central portion of $\mathfrak{G}^2 \subseteq \mathbf{Z}^2$, an example of a rotational forest and jungle.

The operations R and D commute, for if $\mathfrak{V} \subseteq \mathbf{Z}^2$, $D(R(\mathfrak{V})) = (\mathbf{Z}^2 \setminus \mathfrak{C}^2) \cup \{(-2v_2, 2v_1) \mid \underline{v} \in \mathfrak{V}\} = R(\mathbf{Z}^2 \setminus \mathfrak{C}^2) \cup R(\{(2v_1, 2v_2) \mid \underline{v} \in \mathfrak{V}\}) = R[(\mathbf{Z}^2 \setminus \mathfrak{C}^2) \cup 2\mathfrak{V}] = R(D(\mathfrak{V}))$.

Hence any pretty subset $\mathfrak{V}$ of $\mathbf{Z}^2$ is rotational, for if $\mathfrak{V} \sim D(\mathfrak{V})$, $R(\mathfrak{V}) \sim R(D(\mathfrak{V})) = D(R(\mathfrak{V}))$, so $R(\mathfrak{V})$ is pretty. But there are only two equivalence classes of pretty subsets, the members of one being recurrent and not corecurrent, and of the other vice versa. So $\mathfrak{V}$ and $R(\mathfrak{V})$ belong to the same class, i.e. $\mathfrak{V} \sim R(\mathfrak{V})$, so $\mathfrak{V}$ is rotational. In fact we shall see later that the pretty subsets of $\mathbf{Z}^2$ are the only non-elementary ones with this property.

Lemma 20 *(i) If $\mathfrak{V} \subseteq \mathbf{Z}^2$ is rotational, a forest and a jungle then $\exists \mathfrak{U} \subseteq \mathbf{Z}^2$ with $\mathfrak{U} \sim \mathfrak{V}$ and $R(\mathfrak{U}) = \mathfrak{U}$.*
(ii) If $\mathfrak{U} \subseteq \mathbf{Z}^2$ is a forest and a jungle and $R(\mathfrak{U}) = \mathfrak{U}$ then $\mathfrak{U}$ contains $(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$, $(2,1)$, $(1,2)$, $(-1,2)$, $(-2,1)$, $(-2,-1)$, $(-1,-2)$, $(1,-2)$, $(2,-1)$, $(2,0)$, $(0,2)$, $(-2,0)$ *and* $(0,-2)$, *and does not contain* $(1,1)$, $(-1,1)$, $(-1,-1)$, $(1,-1)$, $(2,2)$, $(-2,2)$, $(-2,-2)$ *or* $(2,-2)$.
Proof (i) We have $\mathfrak{V} \sim R(\mathfrak{V})$, so say $\mathfrak{V} = R(\mathfrak{V}) + \underline{v}$. If $v_1 \not\equiv v_2 \pmod 2$, let $\mathfrak{U} = \mathfrak{V} - (\frac{v_1 - v_2 + 1}{2}, \frac{v_1 + v_2 + 1}{2})$. Then $R(\mathfrak{U}) = R(\mathfrak{V}) - (\frac{-v_1 - v_2 - 1}{2}, \frac{v_1 - v_2 + 1}{2}) = \mathfrak{V} - (\frac{v_1 - v_2 - 1}{2}, \frac{v_1 + v_2 + 1}{2}) = \mathfrak{U} + (1,0)$. So $(0,0) \in \mathfrak{U} \iff (1,0) \in \mathfrak{U} \iff (1,1) \in \mathfrak{U} \iff (0,1) \in \mathfrak{U}$. Thus either $\mathbf{Z}^2 \setminus \mathfrak{U}$ is not totally disconnected, contradicting $\mathfrak{U}$ a jungle, or $\mathfrak{U}$ contains a cycle, contradicting $\mathfrak{U}$ a forest. So $v_1 \equiv v_2 \pmod 2$. Let $\mathfrak{U} = \mathfrak{V} - (\frac{v_1 - v_2}{2}, \frac{v_1 + v_2}{2})$. Then $R(\mathfrak{U}) = R(\mathfrak{V}) - (\frac{-v_1 - v_2}{2}, \frac{v_1 - v_2}{2}) = \mathfrak{V} - (\frac{v_1 - v_2}{2}, \frac{v_1 + v_2}{2}) = \mathfrak{U}$.
(ii) This is trivial just by working through the definitions. □

Lemma 21 *If $\mathfrak{V} \subseteq \mathbf{Z}^2$ is grotesque then $\mathfrak{V}$ is not rotational.*
Proof Suppose $\mathfrak{V} \subseteq \mathbf{Z}^2$ is grotesque and rotational. By Lemma 13, $\mathfrak{V}$ is a forest and by a remark above, it is a jungle. So by Lemma 20 (i), we have $\mathfrak{V}_0 \sim \mathfrak{V}$, grotesque, with $R(\mathfrak{V}_0) = \mathfrak{V}_0$. Now by Lemma 20 (ii), $\{(1,0), (2,1), (2,0)\} \subseteq \mathfrak{V}_0$. But since $\mathfrak{V}_0$ is grotesque, $\mathfrak{V}_0 \sim D(\mathfrak{U}_0)$ with $\mathfrak{U}_0$ grotesque, say $\mathfrak{V}_0 = D(\mathfrak{U}_0) + \underline{v}_0$. Now $D(\mathfrak{U}_0) \cap [2\mathbf{Z}_2 + (1,1)] = \emptyset$, so $\{(1,0), (2,1), (2,0)\} \cap [2\mathbf{Z}_2 + (1,1) + \underline{v}_0] = \emptyset$. Thus $\underline{v}_0 \in 2\mathbf{Z}_2$, say $\underline{v}_0 = 2\underline{u}_0$. Then if $\mathfrak{V}_1 = \mathfrak{U}_0 + \underline{u}_0$, $\mathfrak{V}_1$ is grotesque and $\mathfrak{V}_0 = D(\mathfrak{V}_1)$. But $D(R(\mathfrak{V}_1)) = R(D(\mathfrak{V}_1)) = R(\mathfrak{V}_0) =$

$\mathfrak{V}_0 = D(\mathfrak{V}_1)$ so $R(\mathfrak{V}_1) = \mathfrak{V}_1$ and we can repeat the process, whence $\mathfrak{P}^2 \subseteq \mathfrak{V}_0 \subseteq \mathfrak{Q}^2$, so $\mathfrak{V}_0$ is pretty, a contradiction. $\square$

Moreover, Paul Balister and Alex Selby proved:

Theorem 22 *If $\mathfrak{V} \subseteq \mathbf{Z}^2$ is a maximal forest and is periodic, then $\mathfrak{V}$ is not rotational.*

Proof Suppose $\mathfrak{V} \subseteq \mathbf{Z}^2$ is a periodic and rotational maximal forest. By Lemma 8, it is a jungle, so contains an infinite component $\mathfrak{I}$ as it is non-empty. Now $\{\mathfrak{V} - \underline{v} \mid \underline{v} \in \mathfrak{I}\} \subseteq \{\mathfrak{W} \subseteq \mathbf{Z}^2 \mid \mathfrak{W} \sim \mathfrak{V}\}$ which is finite, so $\exists \underline{v}, \underline{w} \in \mathfrak{I}$, $\underline{v} \neq \underline{w}$, with $\mathfrak{V} - \underline{v} = \mathfrak{V} - \underline{w}$. Let $\mathfrak{K}$ be the path in $\mathfrak{V}$ joining $\underline{v}$ and $\underline{w}$, and let $\tau = \max\{\mathrm{d}(\underline{u}, \underline{v}) \mid \underline{u} \in \mathfrak{K}\}$. Then for each $k \in \mathbf{Z}$, $\mathfrak{K} - k(\underline{w} - \underline{v}) \subseteq \mathfrak{V} - k(\underline{w} - \underline{v}) = \mathfrak{V}$ is the path in $\mathfrak{V}$ joining $\underline{v} - k(\underline{w} - \underline{v})$ and $\underline{w} - k(\underline{w} - \underline{v}) = \underline{v} - (k-1)(\underline{w} - \underline{v})$, with $\mathrm{d}(\underline{u}, \underline{v} - k(\underline{w} - \underline{v})) \leq \tau$ for each $\underline{u} \in \mathfrak{K} - k(\underline{w} - \underline{v})$. So $\bigcup_{k \in \mathbf{Z}} (\mathfrak{K} - k(\underline{w} - \underline{v}))$ is an infinite connected subgraph of $\mathfrak{V}$ lying (in $\mathbf{R}^2$) between the lines

$$(w_1 - v_1)y - (w_2 - v_2)x = w_1 v_2 - w_2 v_1 \pm \tau \cdot \mathrm{d}(\underline{v}, \underline{w}).$$

Now $\{\mathfrak{V} + ((w_1 - v_1)\ell, -(w_2 - v_2)\ell) \mid \ell \in \mathbf{Z}\} \subseteq \{\mathfrak{W} \subseteq \mathbf{Z}^2 \mid \mathfrak{W} \sim \mathfrak{V}\}$ which is finite, so $\exists \ell, m \in \mathbf{Z}$, $\ell \neq m$, with $\mathfrak{V} + ((w_1 - v_1)\ell, -(w_2 - v_2)\ell) = \mathfrak{V} + ((w_1 - v_1)m, -(w_2 - v_2)m)$. So if t is an integer multiple of $(\ell - m)$ with $t \cdot \mathrm{d}(\underline{v}, \underline{w}) > 2\tau$, there is an infinite connected subgraph of $\mathfrak{V} + ((w_1 - v_1)t, -(w_2 - v_2)t) = \mathfrak{V}$ lying between the lines $(w_1 - v_1)[y - (w_1 - v_1)t] - (w_2 - v_2)[x + (w_2 - v_2)t] = w_1 v_2 - w_2 v_1 \pm \tau \cdot \mathrm{d}(\underline{v}, \underline{w})$, i.e.

$$(w_1 - v_1)y - (w_2 - v_2)x = w_1 v_2 - w_2 v_1 + [t \cdot \mathrm{d}(\underline{v}, \underline{w}) \pm \tau]\mathrm{d}(\underline{v}, \underline{w}),$$

which form a region disjoint from that between the other two lines. But $\mathfrak{V} \sim R(\mathfrak{V})$, so $\mathfrak{V}$ also contains a pair of infinite connected subgraphs contained within disjoint regions between pairs of lines in a perpendicular direction, so contains a cycle as shown in Figure 8, contradicting $\mathfrak{V}$ being a forest. $\square$

In the light of the previous two results, it was suspected that there were no subsets of $\mathbf{Z}^2$ which were rotational, recurrent and corecurrent maximal forests. Indeed, such a subset must be elementary, so by Lemma 17, we need only look for such a subset $\mathfrak{V}$ with $\mathfrak{C}^2 \not\subseteq \mathfrak{V}$ and $\mathbf{Z}^2 \setminus \mathfrak{C}^2 \not\subseteq \mathfrak{V}$. However, a subset $\mathfrak{R}^2$ with these properties was in fact found by Alec Edgington and Alex Scott. Its central portion is shown in Figure 7. We leave it to the reader to work out its definition and prove that it has the desired properties.

The Venn diagram in Figure 9, known as the gas mask, summarises our results for the case $n = 2$:

Theorem 23 *If $\mathfrak{V} \subseteq \mathbf{Z}^2$ is a maximal forest, it can possess precisely those combinations of the properties of being a jungle, corecurrent, periodic, grotesque, pretty and rotational, which are shown on the gas mask.*

Proof That $\mathfrak{V}$ cannot possess any other combinations of the properties follows from Lemmas 8 (i), 14, 16 and 21, Theorem 22 and various remarks above. The subsets $\mathfrak{J}^2, \mathfrak{G}^2, \mathfrak{M}^2, \mathfrak{S}^2, \mathfrak{P}^2, \mathfrak{Q}^2, \mathfrak{R}^2$ of $\mathbf{Z}^2$ (marked on the gas mask in the appropriate places) show that some of the combinations can arise. The reader should have no difficulty in constructing examples of subsets with the other combinations of properties shown. $\square$

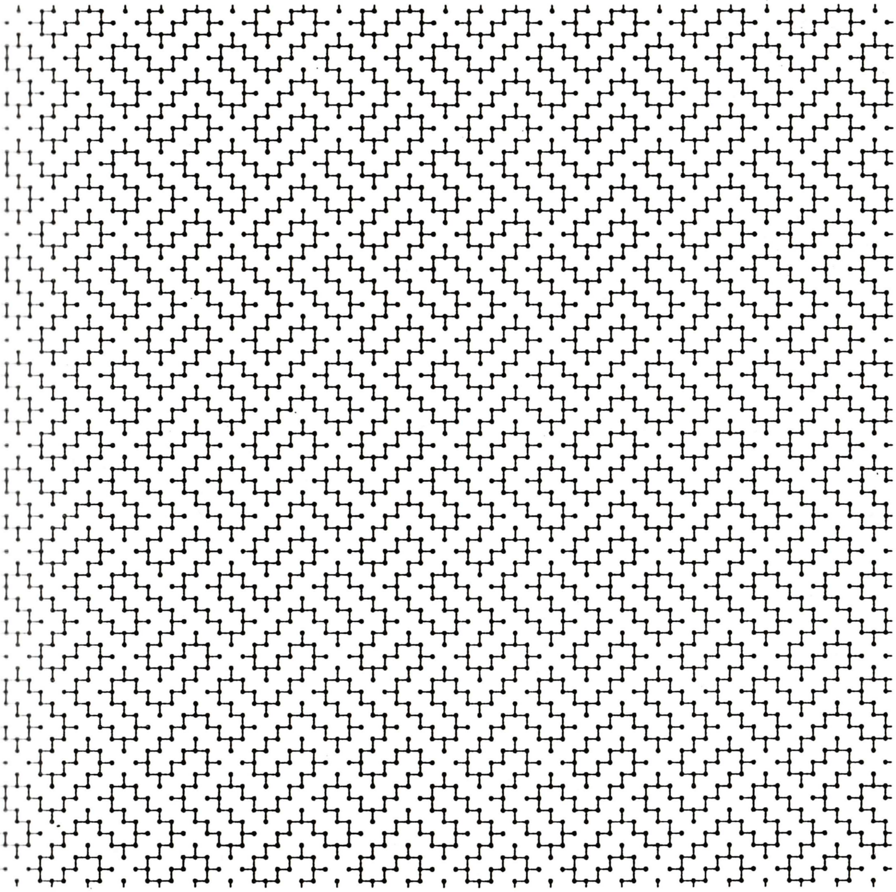


Figure 7 $\mathfrak{R}^2$

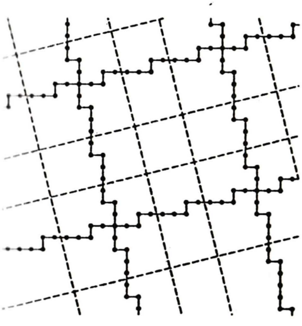


Figure 8

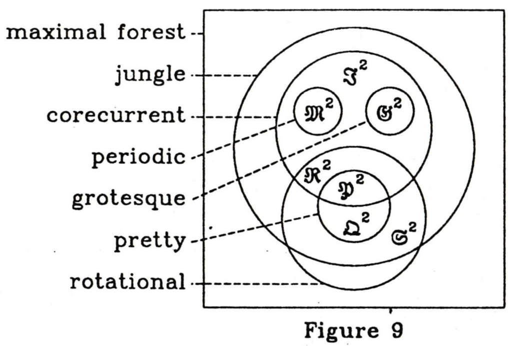


Figure 9

13, 31 and the $3x+1$ Problem

Frazer Jarvis

Problem: Given a positive integer, apply the following algorithm. If it is even, halve it, and if it is odd, multiply it by three, and add one. Repeat this process iteratively.

Example: Start with 3. This is odd, so multiply by three and add one. This gives 10, which is even, so halve this, to obtain 5. Repeated action of this process gives 16, 8, 4, 2, 1, 4, 2, 1,... and we have entered a loop.

Conjecture: Whichever number we start with, the process will always culminate in the loop ..., 4, 2, 1, 4, 2, 1,...

Define: $T(n)$ = the result of the algorithm on n. Define $T^k(n) = T(T^{k-1}(n))$.

The problem dates from before World War II, when it seems to have been proposed by Lothar Collatz, whilst a student at Hamburg [1], but it has resisted all subsequent attempts at solution, despite having been verified up to 10^{12}.

The behaviour of individual numbers of the sequence $n, T(n), T^2(n), \ldots$ is too erratic to discover any useful non-trivial result, so the properties which should be investigated are those which relate to the sequence as a whole. A natural choice is the number of steps k until $T^k(n) = 1$ (assuming the conjecture). So:

Define: $f(n) = \inf\{m \mid T^m(n) = 1\}$.

For example, $f(1) = 0, f(2) = 1, f(3) = 7, f(4) = 2$, etc.; in general, $f(2n) = f(n)+1$, and $f(2n-1) = f(6n-2)+1$ $(n > 1)$. As we tabulate $f(n)$, certain patterns emerge.

An easy one is that for all $K \geq 1$, $f(8K+4) = f(8K+5)$.

To prove this, note that by operation of the algorithm:

$$8K+4 \longrightarrow 4K+2 \longrightarrow 2K+1 \longrightarrow 6K+4$$

$$8K+5 \longrightarrow 24K+16 \longrightarrow 12K+8 \longrightarrow 6K+4$$

and so $T^3(8K+4) = T^3(8K+5)$. So $f(T^3(8K+4)) = f(T^3(8K+5))$. But, by definition, $f(T^k(n)) = f(n) - k$, so $f(8K+4) = f(8K+5)$ as required.

Similarly $f(16K+2) = f(16K+3)$; $f(32K+22) = f(32K+23)$ etc.

In any interval, a few values of $f(n)$ will occur often. For instance, in the range 4900 to 4999, $f(n)$ takes only thirteen distinct values. In this range, $f(n) = 134$ for 29 values of n, $f(n) = 41$ has 72 solutions, and so on.

But the result I would like to investigate is more apparent by looking at differences between consecutive values of $f(n)$. So, with this in mind:

Define: $g(n) = f(n+1) - f(n)$.

So, as a corollary to a previous result, $g(8K+4) = 0$ $(K \geq 1)$.

For what follows next, we need a little elementary number theory.

Theorem *If a, b are coprime, then there exist integers x, y such that $ax + by = 1$.* □

A rigorous proof is easy to find, but to construct x, y, we run the Euclidean algorithm in reverse; e.g. for $a = 12$, $b = 7$,

$$12 = 1 \times 7 + 5 \quad 7 = 1 \times 5 + 2 \quad 5 = 2 \times 2 + 1 \quad 2 = 2 \times 1$$

Then $1 = 5-2\times2 = 5-2\times(7-1\times5) = 3\times5-2\times7 = 3\times(12-1\times7)-2\times7 = 3\times12-5\times7$. So $x = 3$, $y = -5$ solves $12x + 7y = 1$.

An easy corollary is:

Theorem *If d is a multiple of (a, b), the highest common factor of a and b, then there exist integers x, y such that $ax + by = d$. Furthermore, if x_0, y_0 are solutions to this, then so are $x_0 + kb$, $y_0 - ka$ for all integers k, and all solutions are of this form.* □

If we now restrict ourselves to cases where a and b are coprime, we define

$$d_{a,b} = \min\{\sqrt{x^2 + y^2} \mid ax + by = n\}$$

which is well-defined, since the function $f(X) = \sqrt{(x_0 + bX)^2 + (y_0 - aX)^2}$ is convex. Now if X runs through the integers, $\min\{f(X) \mid X \in \mathbf{Z}\}$ is well-defined.

Now we return to the $3X + 1$ problem, and tabulate the values of $g(n)$ for $n = 1, 2, \ldots$ we may observe a rather curious result, namely that $d_{13,31}(g(n))$ is always "small"; in other words, the difference between two consecutive values of $f(n)$ is expressible as $13x + 31y$ for small values of x and y. For example, the values of $f(n)$ are tabulated below for $5385 \leq n \leq 5399$:

n	$f(n)$	$g(n)$	x	y	$[d(g(n))]^2$
5385	147	-80	1	-3	10
5386	67	80	-1	3	10
5387	147	-80	1	-3	10
5388	67	0	0	0	0
5389	67	-39	-3	0	9
5390	28	0	0	0	0
5391	28	88	2	2	8
5392	116	-49	1	-2	5
5393	67	93	0	3	9
5394	160	0	0	0	0
5395	160	-44	-1	-1	2
5396	116	0	0	0	0
5397	116	-49	1	-2	5
5398	67	0	0	0	0
5399	67	49	-1	2	5

where $d(k) = d_{13,31}(k)$.

This phenomenon occurs for all numbers greater than about 20, and as far as 20000 at least, and seems to hold even for random samples of twenty consecutive seven-digit numbers.

For smaller numbers, from 1 to 1000, computer checking has shown that the numbers 13 and 31 perform significantly better than any other pair with a, b coprime and less than 50. [2]

We consider the pair n, $n + 1$. The algorithm essentially consists of a halving or a trebling. If we look at the operation $X \mapsto 3X + 1$, we may neglect the addition of one, because it is fairly insignificant. Then for n to be transformed to 1, we need $a(n)$ halvings and $b(n)$ treblings, with $a(n) + b(n) = f(n)$. Similarly, for $n + 1$ to be transformed to 1, we need $a(n + 1)$ halvings and $b(n + 1)$ treblings, with $a(n + 1) + b(n + 1) = f(n + 1)$.

So, to summarise,

$$n \cdot \frac{3^{b(n)}}{2^{a(n)}} \simeq 1 \quad a(n) + b(n) = f(n)$$

$$(n+1) \cdot \frac{3^{b(n+1)}}{2^{a(n+1)}} \simeq 1 \quad a(n+1) + b(n+1) = f(n+1)$$

For sufficiently large n,

$$\frac{3^{b(n)}}{2^{a(n)}} \simeq \frac{3^{b(n+1)}}{2^{a(n+1)}}$$

Put $a = a(n + 1) - a(n)$, $b = b(n + 1) - b(n)$; then

$$\frac{3^b}{2^a} \simeq 1 \tag{1}$$

Also, $a + b = a(n + 1) - a(n) + b(n + 1) - b(n) = f(n + 1) - f(n) = g(n)$.

$$a + b = g(n) \tag{2}$$

$$a, b \in \mathbf{Z} \tag{3}$$

Solving (1), (2) gives:

$$a \simeq g(n) \cdot \log_6 3$$

$$b \simeq g(n) \cdot \log_6 2$$

Thus

$$\frac{b}{g(n)} \simeq \log_6 2$$

But $b, g(n) \in \mathbf{Z}$, so to find possible values of $g(n)$, look for rational approximations to $\log_6 2$. To find good rational approximations, the obvious method is to look at the continued fraction expression. I give this in standard notation; explanations may be found in any number theory textbook.

$$\log_6 2 = [0, 2, 1, 1, 2, 2, 3, 1, 5, 2, \ldots]$$

giving successive convergents:

$$\frac{0}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{5}, \frac{5}{13}, \frac{12}{31}, \frac{41}{106}, \frac{53}{137}, \frac{306}{791}, \frac{665}{1719}, \ldots$$

The denominators of these fractions are the best possible values of $|g(n)|$. Thus, the best values for $|g(n)|$ are $1, 2, 3, 5, 13, 31, 106, \ldots$ and there are the numbers 13 and 31! Each successive pair $(1, 2); (2, 3); (3, 5); \ldots$ seems to be a "basis" for the values of $g(n)$ for a certain range of n (this occurs because of the simplifications involved in the analysis above).

(a,b)	n for which (a,b) is best basis
$(2,3)$	$n = 1,2$
$(3,5)$	$3 \leq n \leq 8$
$(5,13)$	$9 \leq n \leq 24$
$(13,31)$	$25 \leq n \leq$?

So in fact, the result about 13 and 31 should not surprise us.

To end with, I will leave the interested reader (if there are any) with a few conjectures. I think these are all true, but very difficult to prove!

1). Let $h(n)$ be the highest member of the sequence $n, T(n), T^2(n), \ldots$ It is clear that $\underline{\lim}(\frac{h(n)}{n}) = 1$, by taking $n = 2^k$, and $\overline{\lim}(\frac{h(n)}{n}) = \infty$, by taking $n = 2^k - 1$; then $T^{2k}(n) = 3^k - 1$. Does $\lim(\frac{1}{N}\sum_1^N \frac{h(n)}{n})$ exist?

2). Given $k \in \mathbb{N}$, does there exist n for which $f(n) = f(n+1) = \ldots = f(n+k)$?

3). Set

$$m(n) = \begin{cases} 1 & \text{if f(n)=f(n+1),} \\ 0 & \text{otherwise.} \end{cases}$$

Does $\lim(\frac{1}{N}\sum_1^N m(n))$ exist and if so, what is it?

4). Does $\lim[\frac{1}{N \log N}\sum_1^N f(n)]$ tend to a finite non-zero limit?

The author would be delighted to hear of progress made on any of these problems!

References

[1] Gardner, M., *Wheels, Life and Other Mathematical Amusements.*

[2] Many thanks to John Croft and Graham Nelson for these computer results.

On the Epistemological and Metaphysical Problems of Probability

Alan Stacey

The language of probability theory is common in everyday conversation and simple probabilistic ideas are taught alongside elementary geometry and algebra in classrooms. However, if one enquires of someone what they mean when they talk about the probability of an event they usually have great difficulty in providing any reasonable answer and often freely admit that they do not know. It is my contention that although probability has some meaning (in the context of the actual events rather than an axiomatic system) we are a long way from understanding what it is.

The subtlety of probabilistic ideas is reflected in the age of the subject. Although there has been a very vague notion of probability ever since mankind started to gamble (at least as long as 5500 years ago) precise calculations of probabilities can only be traced back to the sixteenth century and the modern subject of probability is traditionally considered to have started with a correspondence between Fermat and Pascal in 1654 in which they solved various problems concerning simple games of chance. Their ideas were extended by Christianus Huygens, Jakob Bernoulli and others, culminating in the work of Pierre Simon de Laplace (1749-1827). In his *Philosophical Essay on Probability* [3], he explains probability as follows:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favourable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favourable cases and whose denominator is the number of all cases possible.

It is easily seen that this definition is not wholly satisfactory. Keynes referred to the main idea behind this theory as the Principle of Indifference, which states that two events are judged to be equiprobable if there is no reason to believe one more than the other. There is no rule to tell us exactly when we may apply the principle. For instance, it can be used to show that any statement whose truth value we do not know is true with probability half. Even worse, however, is the impossibility of applying the principle in many cases where we might want to compute a probability such as tossing a biased coin. The definition certainly does not allow for experimental evidence to be used to estimate the probability of a head, and there are no equiprobable events in sight, unless we consider both sides still to have equal probability if we do not know how the coin is biased.

It is clear from the above that without a rule to tell us when we may regard events as being equiprobable, the definition of Laplace is dreadfully incomplete. Attempts to define 'equiprobable' lead to further problems in how to tell when we have relevant information which makes two events no longer equiprobable. For instance if we know there are two candidates in an election then if we know that one was born on a Tuesday and the other on a Thursday we need a rule to tell us that this information is irrelevant and that we may still regard the events of each candidate winning as equiprobable. On the other hand we want to be able to tell that knowing that one candidate has been photographed with babies more often than the other is relevant. The definition of Laplace will clearly not do as it stands since at best it can only give us information when we are totally ignorant. Not only is this too restrictive, but the very idea of obtaining precise knowledge from total ignorance has been roundly attacked as being thoroughly unjustified.

So far I have only considered the epistemological problems with the Laplacian theory, i.e. those problems relating to the determination of a probability. We must also consider the metaphysical problems the definition brings: we want to know what a probability means once we have calculated it. As far as the contemporaries of Laplace are concerned the answer is 'not much'. The received view was of a clockwork Newtonian universe in which probability was a human invention to help make rational decisions. Probability merely reflected our lack of knowledge of the world and for someone with sufficient knowledge of an event the probability would either be 0 or 1. If someone is about to draw a ball from a bag 'at random' there is no such thing as the 'correct probability' that a particular ball will be drawn other than 0 or 1 since a sufficiently knowledgeable person could predict the outcome with certainty.

In practice Laplace got round some of the difficulties of his definition by cheating and ignoring it. In composing actuarial tables Bernoulli and Laplace used the empirical evidence of mortality figures and at one stage Laplace spoke of probability as representing a 'degree of belief'. In so doing they foreshadowed relative frequency theories and subjectivist theories respectively. I shall give a brief description of these more modern types of theory together with a description of a priori theories which are more sophisticated successors of classical Laplacian theories. (This classification of theories is due to Weatherford [7].)

A priori theories are in certain respects refinements of the classical theory of probability and their earliest major proponent was John Maynard Keynes, who earned his fellowship at King's with a dissertation which was later published in an expanded form entitled *A Treatise on Probability* [2]. He argued that probability is a logical relation between two sentences one of which is a proposition and the other of which is to be thought of as evidence. This relation can be determined a priori, not by empirical means. The exact nature of this relation is not easy to understand; in fact Rudolf Carnap, another major advocate of a priori theories, spent over one hundred pages in his *Logical Foundations of Probability* [0] just establishing the languages and formalisations required for the task. Only in certain cases, according to Keynes, can the relation be assigned a numerical value. In those cases Keynes calls upon the 'Principle of Indifference', although rather modified from the form described by Laplace.

Keynes emphasised that probability is always relative to evidence [2]:

> The terms 'certain' and 'probable' describe the various degrees of rational belief about a proposition which different amounts of knowledge authorise us to entertain. All propositions are true or false, but the knowledge we have of them depends on our circumstances; and while it is often convenient to speak of propositions as certain or probable, this expresses strictly a relationship in which they stand to a corpus of knowledge, actual or hypothetical, and not a characteristic of the propositions themselves. A proposition is capable at the same time of varying degrees of this relationship, depending upon the knowledge to which it is related, so that it is without significance to call a proposition probable unless we specify

the knowledge to which we are relating it.

One of the great advantages of a priori theories is that they deal rather neatly with the problem of there being more than one 'real probability'. Consider a situation in which we have two bags. One contains two red balls and one white; the other contains two white balls and one red. Marcus has selected a bag and is about to choose a ball blindly from it. He does not know which bag he has chosen and if he were to assign a probability to his selecting a red ball it would be one-half. Helena knows, however, that he has chosen the former bag so assigns a probability of two-thirds to selecting a red ball. Dilip the Amazing, however, can tell from the state of the universe that Marcus is certain to pick a red ball so assigns a probability of one. Whereas certain theories of probability have difficulty in accomodating these different values as all being correct, the a priori theorist simply points out that they are measuring a proposition relative to different pieces of evidence: in essence, Marcus, Helena and Dilip are measuring completely different things. In this case, for instance, Marcus is working with the relation between 'Marcus is about to draw a red ball' and a sentence telling him the composition of the two bags and the fact that he is about to draw precisely one ball from one of them.

It is important to realise that a priori probabilities exist regardless of the state of the physical world, so any of the above probabilities could have been calculated equally well by any one of the three people provided that they regarded it as relative to a particular piece of evidence. This is, however, one of the main complaints about this sort of theory. We would like to have a probability which is related to our intuitive notion of how often things will happen and not to an abstract relation between sentences. This is the advantage of defining the probability of an event to be the relative frequency of the event as the number of trials tends to infinity.

The relative frequency theory was developed by Richard von Mises and has been applied with great success in many branches of science where a large number of experiments are done (or a large sample is examined) to estimate probabilities. Von Mises maintained, moreover, that a probability value is a measure of an empirical, objective fact about the real world, and that such values are uniquely determined by the state of the world and are not relative to a particular piece of evidence. Probabilities are not assigned to particular one-off events (other than possibly 0 or 1) but are only assigned to special classes of events. According to von Mises [5]:

> The rational concept of probability, which is the only basis of probability calculus, applies only to problems in which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time. Using the language of physics, we may say that in order to apply the theory of probability we must have a practically unlimited sequence of uniform observations.

Many opponents of relative frequency theories regard this restriction as a serious disadvantage. Moreover, this causes a whole new class of problems of deciding precisely what events one can assign probabilities to.

However, there is a much more serious problem with relative frequency theories than their being restricted to only certain types of events. The definition provides probabilities that are unknowable, unverifiable and cannot be known to exist. This is for the simple reason that human experiments, on which the whole theory relies, cannot measure infinitely many events so it cannot be known whether convergence is occurring and, even worse, no putative value (provided it lies in [0,1]) can be disproved. I have heard eminent probabilists criticise philosophers on the grounds that although we may not be able to describe exactly what we mean for a coin to have probability half of landing heads, the Strong Law of Large Numbers (which the 'stupid' philosophers are supposed not to know) tells us exactly what will happen. Unfortunately even if we assume that a probability-one event always happens we do not have infinitely many coin tosses to apply the result to.

The last major type of theory which I shall mention is the subjectivist theory, which was initially developed by Frank P. Ramsey [6] in 1931, and then extended by Bruno de Finetti [1]. The idea behind this is that probability is not an objective fact about the universe but measures the degree of belief held by an individual in a certain proposition. This degree of belief can be tested by certain psychological tests such as the willingness of an individual to place bets on the proposition. Subjectivists also normally insist that a person's beliefs are coherent so that, for example, the sum of probabilities of events of which precisely one must happen should be one. Provided this is the case, however, people are free to choose thr probabilities as they like and no one set of choices is more 'rational' than another.

Perhaps surprisingly, the application of the idea of subjective probability, in the form of Bayesian statistics, has had remarkable success. We are all familiar with Bayes's rule that if a probability space has a partition into events $H_0, ..., H_n$, each with non-zero probability, and E is any event, then:

$$\mathsf{P}(H_k \mid E) = \frac{\mathsf{P}(E \mid H_k) \cdot \mathsf{P}(H_k)}{\sum_{i=0}^{n} \mathsf{P}(E \mid H_i) \cdot \mathsf{P}(H_i)}$$

To quote from a standard textbook on statistics [4]:

> The basic assumption in the solution of the decision problem termed 'Bayesian' is that one's beliefs about nature can be represented by the model of a probability distribution. This distribution has been called a prior distribution, to suggest that it represents the statistician's beliefs as they exist prior to the collection and assimilation of any data. ... The notion of using Bayes's Theorem to modify prior beliefs by incorporating the information from a sample to obtain better educated beliefs is the essence of 'Bayesian' statistics.

Although this method has been greatly attacked for the lack of justification in choosing the initial probabilities, it has had great success in practice since, provided the values chosen are all non-zero, it does not much matter what the initial probabilities are for their effect is quickly swamped by the accumulation of objective data. The subjective theories, unlike most others, justify these arbitrary choices.

Subjective theories also have the advantage over relative frequency theories that they have meaning for individual events and do not need an infinite number of experiments for probabilities to be determined. However the problem with them is, I claim, obvious. We know from experience that a gambler who believes that (six-sided ordinary) dice come up a six with probability one-sixth fares much better than one who believes the probability to be one-half. The Laplacian and a priori theorists have no difficulty in pointing out where the latter gambler is going wrong, and the relative frequency theory will tell us after we perform infinitely many experiments, but the subjectivists tell us that both gamblers are equally rational even though the latter is going rapidly bankrupt. A similar application of this idea to probabilities in quantum mechanics will tell us the rather worrying fact that the tiny probability of the Trinity Great Court fountain jumping a mile into the air will become much larger if our beliefs change; or rather, we feel it ought to be worrying but if we remain strictly within the bounds of subjectivism it is not.

If subjectivism is not the correct view of probability, then I think it still has interest in its own right for the degrees of belief of individuals in particular propositions - if not probability - are certainly important.

Incidentally, a similar argument to the above shows why it would be absurd to claim that probability has no meaning at all. If this were the case it would be impossible to explain why poker players with a sound knowledge of probability do much better than those without and why quantum mechanics has achieved anything at all.

Having just scratched the surface of subject of the interpretation of probability, I hope I have indicated that the matter is both non-trivial and important. The work of probabilists can go on regardless of this for their work is both beautiful and important for solving non-probabilistic problems. However, attempts to define probability in the 'real world' have been of great importance in both the physical sciences and general statistics, as I have already pointed out. I have heard probabilists say that 'philosophers are pompous whereas we [probabilists] are precise', but if we are to use 'precise' probabilities to describe the world around us, it would be nice to know what they mean.

References

[0] Carnap, R., *Logical Foundations of Probability*, University of Chicago Press, 1950.

[1] Finetti, B. de, *Teoria della Probabilitá*, 1970, (translated as *Theory of Probability: A Critical Introductory Treatment*, New York, Wiley 1976).

[2] Keynes, John Maynard, *A Treatise on Probability*, London, Macmillan, 1921.

[3] Laplace, P. S. de, *A Philosophical Essay on Probability*, New York, Dover, 1951 (originally published in 1820).

[4] Lindgren, B. W., *Statistical Theory*, New York, Macmillan, 1976.

[5] Mises, R. von, *Wahrscheinlichkeit, Statistik und Wahrhelt, Einführung in die neue Wahrschein Lichkeitslehre und ihre Anwendung*, Vienna 1928 (translated as *Probability, Statistics and Truth*, London, Macmillan 1957).

[6] Ramsey, F. P., *The Foundations of Mathematics and Other Logical Essays*, London, Kegan Paul 1931.

[7] Weatherford, Roy, *Philosophical Foundations of Probability Theory*, Routledge and Kegan Paul, 1982.

On Smarties and Time Travel
or
Why you haven't got a time machine

William Hurwood

This is the text of a paper which I presented to the New Pythagorean Mathematical Society in Michaelmas 1988. Purists among you who were lucky enough to see the original talk may notice that it has been rewritten in places. (To be more exact it has now been written.)

In this discussion we will take it for granted that I possess a time machine (which I will refer to as a TARDIS from now on.) The exact method by which it works is of no interest to us at all and I will leave it to the engineers and physicists to worry about that. What we are interested in is what we can deduce about the universe following my "borrowing" of a TARDIS.

The main objection against any time travel has always been the paradoxes that it appears to involve. Now, as any set theoreticians will tell you, a theory which permits paradoxes is basically a wrong theory. So in the real universe paradoxes do not occur. This leads us to the First Law of Time, namely:

$$\boxed{\text{Nature abhors a paradox!}}$$

The problem here is what I mean by a time paradox. There are two main types of paradox:

(i) "Going back in time and killing your ancestor."

(ii) "Going back in time and marrying your ancestor."

These are quite different. In the first case you take action to, absolutely and finally, prevent an event occurring which you know must happen, namely your own existence in this case. Once you have done this there is no way out of the paradox. Either you exist or you didn't. This is the sort of paradox prevented by the first law.

The best way to visualise this is to see the whole of space and time as a monstrously complicated system which has already been laid out in its entirety. Every event only happens once and in one way. Thus "altering" the past or the future is impossible. Type (i) paradoxes simply cannot happen. There are no alternate time lines to consider. Of course this way of looking at things does rather spoil the concept of free will. But most people haven't got a TARDIS and so are utterly unable to perceive how their choices are predetermined. A difference which makes no difference is no difference.

However, in the second case your action, although bizarre, is not paradoxical. Being your own ancestor doesn't actually prevent the drawing up of a consistent table of where everything was at any particular time. So you could marry (and hence be) your own ancestor.

Suppose however that after a particularly hard day (say, for example, doing a Representation Theory example sheet), you became a little fed up and decided to (suppose) assassinate Augustus anyway and let reality look after itself. Now one of two things could happen. Either you:-

(i) Break the First Law and thus cause the premature end of the universe, or

(ii) Fail to assassinate Augustus.

As is usual in these things, it is possible to succeed in killing Augustus but not cause a paradox. You could do this by, for example, killing him on the day he is known to have died, or possibly by replacing him with an identical actor.

To settle what actually happens we need a Second Law. This is obtained by an extension of the Anthropic Principle:-

$\boxed{\text{The Universe doesn't end.}}$

This is reached by the argument that if I am wrong, no one is ever going to be able to say to me "I told you so, William". Thus I can experiment with my TARDIS in perfect safety. So the second possibility must occur. No matter how determined I am in my grisly task, I am unable to perform it. To a hypothetical observer it would appear that reality conspires to preserve the life of Augustus.

I shall now give a few examples of what I mean, from that greatest of programmes - Doctor Who. In "City of Death" [1], an alien, after being blown through time by an explosion, attempts to travel back in time to warn himself before the event. In the end he does get back in time, but is unable to warn himself. We now see that his failure to do so is inevitable. Another fine example is given in "The Aztecs" [2]. Barbara, having been mistaken for a Goddess (for complicated reasons) attempts to alter history by persuading the Aztecs to give up human sacrifice, so that they won't be massacred by the Spanish invaders. The harder she tries, the more she finds obstacles on her path and she is eventually forced to give up. Even the Doctor is not immune from this effect. In "Genesis of the Daleks" [3], he attempts to destroy the Dalek race before it became established. All he is able to do is to entomb the embryonic Daleks for a thousand years. In fact the Daleks always had been entombed at the time, and the Doctor was simply previously unaware of this fact.

An example of the "conspiracy" can be found in "Mawdryn Undead" [4], where the Brigadier meets himself. In order to prevent the future Brigadier from warning the past one about what will happen to him, "reality" inflicts a nervous breakdown on the past Brigadier so that he completely forgets everything that happens to him.

Of course, Doctor Who is only fiction and occasionally gets it wrong. This is most notable in "Day of the Daleks" [5], where the whole plot revolves around an "alternative" future, something we now see to be a meaningless statement.

You may be wondering by now what Smarties have got to do with all this. They serve is an illustration of some simple consequences of the Laws of Time as outlined above. Even if you haven't got access to a TARDIS, you should be able to persuade a friend to send a packet of Smarties into the past for you.

Try the following experiment. You have a packet of Smarties in front of you in two different places. One is the "young" packet which you bought from Sainsbury's, the other is the "old" packet which your friend gave you. At the end of the evening, you will give the "young" packet to your friend to send through time and she will return it to you yesterday morning. (If you can't find anyone else, I would be very happy to transmit any packets of Smarties that anyone should give me. †)

---

† The author demonstrated this process during the talk with two packets of Smarties,

Now you can do anything you like to the "old" packet since you know nothing at all about its future. But the "young" packet is protected by reality, on pain of the untimely demise of the universe. So you could eat a single sweet from the "young" packet, as you are unable to tell from the "old" packet that this has been done. More interestingly, however, you could mix all the Smarties in a bowl, and then with your eyes shut eat half of them. Assuming that the universe doesn't end, you must have eaten precisely those Smarties which came from the "young" packet. So the "conspiracy" is able to do some pretty mind-numbing things to probability if it has to. More formally, let us say that an event Q is necessary if it must occur to prevent the end of the Universe. Clearly then, by simply increasing the number of Smarties in a packet, we find

$$\forall \epsilon > 0 \quad \exists Q \quad \text{necessary} : \mathcal{P}(Q) < \epsilon.$$

An unsolved problem in temporal probabilistics concerns events R with $\mathcal{P}(R) = 0$, but such that R is not logically prevented from happening: e.g. $R =$ 'A randomly chosen real number from [0,1] is rational'. Can such an R be made necessary?

From this, I must emphasize the importance of the observer. Suppose on receiving the "old" packet you had immediately weighed it and found that it weighed exactly the same as the original packet. Then you couldn't eat even one Smartie from the "young" packet. So we find that what is important is not what we might have done, but what we actually did.

Finally, this explains why those among you who may be sceptical, may have been unable to get your sticky paws on a TARDIS. All you want to do is to try to prove me wrong by assassinating Augustus, or eating the wrong Smartie. Reality would have to bend probablity no end to get that sorted out. It's much easier to simply prevent you from getting a TARDIS, so you can do no damage. This leads us directly on to the Third Law of Time:-

$\boxed{\text{You haven't got a TARDIS.}}$

Although I have.

References

[1] Serial 5H, Douglas Adams, 1979.

[2] Serial F, John Lucarotti, 1963 (the four episodes "The Temple of Evil", "The Warriors of Death", "The Bride of Sacrifice", "The Day of Darkness", now usually together referred to as "The Aztecs").

[3] Serial 4E, Terry Nation, 1976.

[4] Serial 6F, Peter Grimwade, 1983.

[5] Serial KKK, Louis Marks, 1973. †

proving them to be the same by showing that the tops of both were green W's - Ed.

† In fact, the plot can be resolved consistently by invoking the parallel universes model hinted at in "Inferno" (Serial DDD, 1970). There is no need to resort to counter-intuitive notions such as Doctor Who being merely fictional - Ed.

Problems Drive 1989

Set by Marcus Moore and Alan Stacey

Twelve questions must be answered in one hour and ten minutes, with pairs playing as teams. To complicate things, the teams can only see each question for five minutes before they have to pass it on. Under such circumstances, the teams display remarkable creativity and so a prize of a clockwork hedgehog is awarded for the most substantially wrong answer; also available are a bottle of port and the obligation to set next year's problems for the winners, and a wooden spoon for the losers.

0. St. Peter is guarding the gates of Heaven when he is approached by three creatures each seeking admittance. He has been informed in advance by God (who is omniscient, never lies, but isn't always as helpful as St. Peter might want) that one of these creatures is a saint (who always tells the truth) and is to be admitted, one is the Devil (who is thoroughly evil and always lies) who is to be sent back to Hell and the other is a mortal (with a strange disorder causing him to make only true statements on alternate days and make only false statements on the remaining (alternate) days) who is to be sent to Purgatory.

The only difference in outward appearance of these creatures is that they carry different colour shields. Denoting by G, R and B the creatures carrying a green, red and blue shield respectively they make the following statements:

G: I am a mortal. R is the Devil.

R: G is the devil.

B: I am a saint.

Deciding (correctly) that he cannot tell between them, St. Peter sends them all away. During the night however, one of the creatures (fearing discovery) swaps his shield with that of someone else. God tells St. Peter that precisely two shields have been swapped, but refuses to say any more.

The next day (so in particular the mortal is on a different parity) they return, and denoting by $G^\star$, $R^\star$ and $B^\star$ the creatures now holding the shields coloured green, red and blue respectively, they make the following statements:

$G^\star$: I was carrying a different shield yesterday.

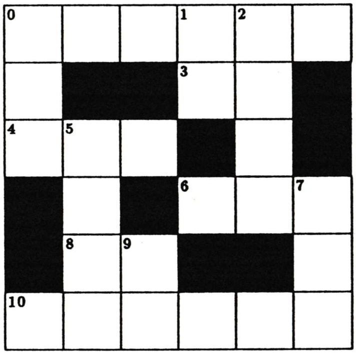
$B^\star$: The person who was bearing a blue shield yesterday (i.e. B) was a saint.

St. Peter (being a perfect logician) deduces who is who. In terms of $G^\star$, $R^\star$ and $B^\star$, whom does he tell to "go to Hell", whom does he admit, and whom does he dispatch to Purgatory? Also, which shields were swapped?

1. What is the smallest $n > 2$ such that

$$2^{3^{\cdot^{\cdot^{\cdot^{(n-2)^{(n-1)^{n}}}}}}} > (10n)^{(10(n-1))^{\cdot^{\cdot^{\cdot^{40^{30^{20}}}}}}}\ ?$$

2. Solve the following crossnumber:



All answers are in base n for some positive integer $n \geq 2$ to be determined. All clues are in base ten. All numbers are written without leading zeros.

Across
0. $(n+2)!$
3. A prime
4. A cube
6. (1 down)$\times$(3 across)
8. Divisible by three
10. $n\times$(A product of twin primes)

Down
0. A cube
1. A prime
2. A square
5. (0 down)$\times$(3 across)
7. A Fibonacci Number
9. Divisible by two

3. Black and White play a game on an 8×3 board, with columns designated A, B and C, and rows numbered 0-7. Black has N pieces, which she places on distinct squares of her choice in rows 0, 1 and/or 2 at the start of the game. Thus $N \leq 10$. White has 1 piece, which starts on B7.

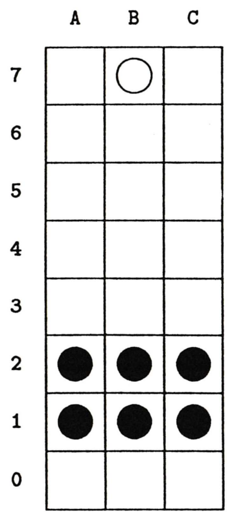
White moves by moving her piece exactly one square diagonally into an empty square.

Black moves by pushing exactly one of her pieces one square forward into an empty square.

White makes the initial move, and turns alternate thereafter.

The game ends when one player cannot make a move on her turn, and that player has then lost. Thus Black must trap White before all of Black's pieces pile up on the back row.

What is the least N such that Black can force a win? For this N, what starting position might Black choose?



4. An elite dining society has nine members and an octagonal dining table with eight chairs spaced equally around it. Each night that the society meets precisely eight members are present and they sit at the table in the usual way. Every member is so obnoxious that once member X has dined next to member Y (s)he refuses to dine with Y ever again (for all X, Y, distinct members). What is the maximum number of dinners, M, the society can hold in its history?

Denoting the members by the letters A-I, give M seating plans which enable M dinners to happen.

5. ...and so the Fairy Godmother took the Princess to a lily pond. Seated on a lily pad were three identical amphibia A, B and C. The fairy said:

"One of these three is in fact the handsome prince you asked me about. The two others are either a frog and a toad, both frogs, or both toads. In turn, each of the three will state the type (prince/frog/toad) of each of the other two. For example A might say "B is a frog. C is the prince." Toads only make true statements and frogs only make false statements, but the prince can make either. If you can deduce which is the prince, you may marry him."

In fact, B was the prince, and A and C were both frogs. B was already engaged to a toad, and resented the fairy's matchmaking. He listened to A's statements, and then made the **unique** pair of statements that guaranteed, irrespective of what C might say, that the princess could not identify him.

After hearing the six statements, the princess deduced that A must be a frog. She could not tell which of B and C was the prince, and wrongly guessed C. Fortunately, she came to like C, and they lived happily ever after.

What statements did A and B make?

6. Match the christian names and surnames of the mathematicians below:

Christian Names (and titles)	*Surnames*
Andrey Nikdayevich	Cantor
Baron Augustin Louis	Cauchy
Bernhard	Descartes
Karl Friedrich	Euler
Emmy	Fermat
Évariste	Galois
Georg	Gauss
George	Green
Sir Isaac	Kolmogorov
Karl	Newton
Leonhard	Noether
Pierre de	Riemann
René	Weierstrass

7. In the following addition sum, each letter stands for a digit, and no two letters stand for the same digit. What are **QARCH** and **EUREKA**?

```
 QARCH
 QARCH
 QARCH
 QARCH
 QARCH
 QARCH
------
EUREKA
```

8. In the distant land of Archimedia the basic unit of currency is the Eureka and there are five Qarchs to the Eureka.

A gambler in the casino of this land is playing a game at which on each round he may bet any non-negative amount of money up to the total amount he has provided he bets an exact number of Qarchs. With probability one-half he loses his stake. Otherwise he wins an amount equal to his stake.

He starts the game with one Eureka and he plays three rounds in succession with the aim of maximising his chance of finishing with at least n Qarchs (where n is a natural number or zero). Find his chance of success for each value of n.

9. At the annual "cattle market", the five college mathematics societies argue over which society is to invite each of five star speakers to give talks. For the sake of anonymity, we shall refer to the speakers as Drs. A, B, C, D and E. Each society makes a series of demands.

Adams:	We must have B if the Quintics have E.
	If the New Pythagoreans get C, we won't take A.
New Pythagoreans:	The Adams can't have C unless we get D.
	A shan't go to the Tensors.
Quintics:	We will have E unless the TMS get him.
	If the Tensors have B, we won't have A.
Tensors:	If the New Pythagoreans get A, we want E.
	If we can't have C, neither can the Quintics.
TMS:	Either give us A, or the Tensors can't have C.
	The Adams can't have B if we have A.

Allocate exactly one speaker to each society in such a way as to satisfy all their requirements.

10. Find the next two numbers in each of the following sequences:

(0) $6, 10, 14, 15, 21, \ldots$

(1) $1, 1, 2, 6, 10, 50, \ldots$

(2) $0, 1, 1, 2, 1, 2, 1, 3, \ldots$

(3) $26, 15, 31, 19, 11, 3, 16, 7, \ldots$

11. Ethel is a Platonic solid with centre O and a triangular face PQR. S is the unique point outside Ethel such that $PQRS$ is a regular tetrahedron. In the cases where Ethel is (a) an icosahedron, (b) an octahedron, (c) a tetrahedron determine whether OQS is less than, equal to, or greater than a right angle.

Mathematical Knowledge and Mathematical Understanding

Professor Rolph Schwarzenberger
Mathematics Education Research Centre
University of Warwick

This paper is taken from an unpublished collection of papers in honour of Professor Christopher Zeeman, presented to him upon the occasion of his leaving the University of Warwick to become Master of Hertford College, Oxford. It is partly based upon two talks given by Professor Schwarzenberger to the Society.

On moving from mathematics to mathematics education in 1979, I was struck by the differences. Mathematics depends, at any given time, on a relatively small number of theories; there are clear theorems and there is (usually) a consensus on whether or not their proofs are correct. Mathematics education has too many theories; there are few theorems and even more doubt as to whether alleged results are valid. Good work in mathematics often shows that some result or theory is simpler than was previously supposed. Good work in mathematics education usually shows that the conditions for successful teaching and learning of mathematics are more complex than was previously supposed.

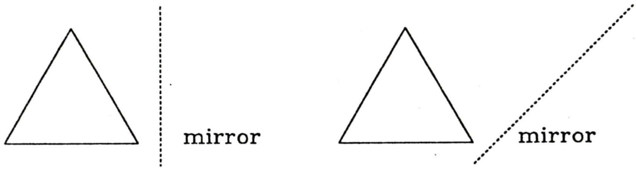
There is also another difference. In mathematics it is assumed to be easy to check what someone knows and, amongst themselves, mathematicians can tell instantly whether or not someone "understands". It must be remembered that mathematicians form a very small subset of the larger set of those who learn or use mathematics and it is often mistakenly assumed, by politicians as well as mathematicians, that it is just as easy to check what a non-mathematician knows and understands. In mathematics education it is accepted that most pupils are not "mathematicians" by inclination or ability. There is a considerable body of research on the difficulty of assessing what a pupil knows, and there is much uncertainty about what precisely "understanding" means. Since the school mathematics curriculum for England and Wales is currently being reformulated in terms of "attainment targets" - by which the government means "clearly specified objectives for what pupils should be know, understand and be able to do" [8] - this uncertainty needs to be brought out into the open.

1. Difficulties of assessing mathematical knowledge and understanding

Research in mathematics education is full of examples which contradict our naive everyday experience that mathematical knowledge and understanding are easy to assess. There seem to be three underlying reasons.

Firstly, it is a common experience that we can hold in our minds - as it were in different compartments - contradictory notions about the same mathematical object. Thus it is a common belief that 0.999... must be different from 1, even among people who are quite sure that 0.333... is equal to 1/3. Students move unconsciously between two different functions called "sin", sometimes fudging the issue by writing sin x^0 or sin x^r, the first being used in trigonometry and the second being used in calculus. Even if notions are not contradictory there is frequent need to think of the same objects in different ways: for example as a number or as a transformation, as a function or as a graph, as a function or as a single element in the set of all functions, and all on. Skemp [23] has called this the varifocal nature of mathematics; it is part of the experience of all mathematicians but starts with the earliest experience of space and number. It means that some doubt must always attach to simplistic attempts by assessors to find out what notions a student holds.

The second reason why mathematical knowledge and understanding are difficult to assess is closely related. Even if the assessor is searching for notions which the student does indeed hold, it may happen that the context of the assessment acts as a distractor. For example, a student may have a fairly good understanding of reflection, and be able to reflect the triangle shown in Figure 1, and yet be unable to reflect the triangle shown in Figure 2. The difference is not in the triangles but in the mirror line.



Figures 1 and 2

A strikingly similar example has been found in research by Herschkowitz and colleagues in Israel [14]. They tested 5th to 8th grade pupils, and also initial teacher-training students and serving teachers, on their ability to recognise the right-angled triangles among a collection of ten different triangles drawn at different angles to the horizontal. There was high success at recognising triangle (a) in Figure 3 as right-angled, and recognition of (b) improved steadily with age, but less than 50% of pupils, and less than 75% of teachers could recognise (c). Possible reasons are outlined in [20].

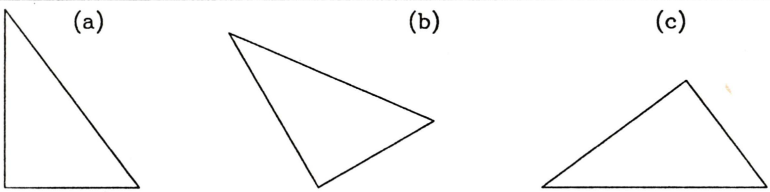


Figure 3

Similarly in work with numbers there is usually a context, which may be unspoken, which determines whether it is appropriate to use natural numbers, integers, real numbers, complex numbers, or more sophisticated values; mathematics, as well as mathematical understanding, often advances by shifting the boundaries between such categories. Therefore any associated assessment is measuring not just a calculation but the appropriateness of the calculation to the context provided.

The third reason why mathematical understanding, in particular, is difficult to assess is that it has features personal to the individual being assessed. It is tempting for an assessor to believe that, because mathematical knowledge is "objective", there is a unique way in which a given piece of knowledge should be understood. Attempts at assessment which have such preconceptions may well miss the real understanding which exists. Even the widely accepted assertion that mathematics is hierarchical is not true without qualification. For example, an empirical study by Denvir, Brown and Eve [7], carried out in connection with the DES-sponsored feasibility study on attainment targets, found that children given word problems which involved division did display a range of strategies that were partially ordered (in the sense that if $S < T$, and a child displayed strategy T, then the child also displayed strategy S). However the strategies were not ordered, nor could they be placed on a set of levels for which it can be said that a strategy on level n requires all strategies on level $n-1$. Previous work by Denvir [5] showed that it was possible only to organise strategies into levels by taking a broader definition: if a level is judged to be attained when two-thirds of the strategies at the level have been displayed, then it was true that no child attained a level without attaining all previous levels. The above results are all concerned with elementary school mathematics but it seems likely that the same is true at all stages: mathematics is hierarchical only in the rough and ready sense of a partial ordering, and there is no uniquely determined pathway through the concepts and strategies involved.

These three underlying reasons why assessment of mathematical knowledge and understanding is difficult can be confirmed at all stages of mathematical development. They apply both to the diagnostic assessment of single individuals and to written or oral class assessment. Clearly with individual assessment there is more chance that the assessor will be sensitive to the personal constructs of the person assessed. A recent study of Denvir and Brown [6] used assessment instruments for 7-11 year old children on number concepts and skills, comparing the results of a class administered assessment with those of an individual diagnostic interview. It found that the results ("pass" or "fail") differed for approximately 1 in 6 cases. The conclusion was that class administered assessment did not accurately measure the achievement of all the children in the class but gave a useful first approximation from which those needing further diagnostic assessment could be identified.

2. Attempts to define mathematical understanding

It is not surprising, given the contrast between our strong conviction that mathematical understanding has a well-defined existence and the difficulty we experience in any assessment of understanding, that many attempts have been made to analyse the different kinds of mathematical understanding which can be observed. Here are some examples:

Instrumental understanding is the ability to apply appropriately remembered rules to the solution of a problem. Instrumental understanding may be displayed whether or not it is accompanied by any knowledge about why the rules work [19]. Originally this kind of understanding was isolated by Skemp as a contrast to other more desirable kinds of understanding; he gave as an example the pupil who knows that the area of a rectangle is obtained by multiplying width times height but does not know why. Vinner [26] has pointed out that this is actually the most common understanding among mathematics students from primary level to college level. He analyses it more closely and distinguishes true instrumental performance from "quasi-instrumental" performance in which, for example, the student knows merely that in sums about areas it is necessary to multiply two numbers together but may nonetheless often obtain the correct answer.

Relational understanding is the ability to deduce specific rules or procedures from more general mathematical relationships. Originally this kind of understanding was seen by many authors as obviously the most desirable [21] on the ground that knowledge of general mathematical relationships - and in particular links between one concept and another - allows the creation of mental schemes which are adaptable to different problems in different contexts. On the other hand it became clear that to accord such primacy to relational understanding is to ignore the importance in mathematics of going "into automatic" and allowing the formal symbolism to take over. Hence the proliferation of further kinds of understanding:

Intuitive understanding is the ability to solve a problem without prior analysis. Its inclusion in this list draws attention to the fact that understanding may exist, and may be in excess of one's formal knowledge, without being articulate or even conscious [3].

Logical understanding is the ability to demonstrate that what has been stated follows by a chain of inferences from what went before [3].

Symbolic understanding is the ability to connect mathematical symbolism and notation with relevant mathematical ideas [24].

It is hard to avoid the feeling that, the more different facets of understanding are isolated and analysed, the further away is the goal of achieving an understanding of mathematical understanding. Subsequent attempts to clarify the different kinds of understanding by distinguishing different modes of thinking (e.g. reflective, intuitive) [22], by distinguishing different sorts of subject matter [11], or by establishing levels of understanding which form the basis of a constructivist model [1], [2], [12], [13], are all subject to the same criticism [18], [19].

In any case, empirical work has demonstrated the impossibility of assigning a pupil's understanding unambiguously to a particular category. Pirie [17] gives an example in which a pupil's understanding of division of fractions would have been categorised differently by observers with access to different episodes of the pupil's behaviour. Even descriptions in terms of levels are not unambiguous: pupils can operate on more than one level simultaneously, or can make sudden shifts from one level to another as a result of changes of context.

On any model of understanding there are at least three important components: items of knowledge, which may well be held in different compartments; conceptual links between items of knowledge; and the ability to select the appropriate knowledge or skills for a given problem. Tall [25] and Haylock [9] have emphasised the importance of regarding understanding not as a static "state" but as a dynamic phenomenon. This approach seems to fit the evidence better than a classification into different kinds of understanding. When we describe our own understanding of a mathematical topic we may well describe a conceptual "state", but when exploring the understanding of another person we become aware of understanding as a description of movement: between one concept and another, linking concepts, adapting to new contexts, testing against new conjectures and so on.

We are therefore led to an alternative view of mathematical understanding which does not merely describe the components of knoweledge and skills which make it up. Knowledge and skills are typically unstable and fragmented; this is why they are difficult to access. Understanding is perhaps the name which we give to the existence of a sufficiently large connected area of stable knowledge together with the ability to traverse this area mentally making connections and selecting procedures.

3. Assessment of understanding

If we view understanding as a sufficiently large, and easily traversed, area of stable knowledge then what are the implications for attempts to assess understanding? Such assessment is the aim of most university and college mathematics examinations. The assumption that understanding, as well as knowledge and skills, can be assessed is often made without question. For example, the government announcement in July 1987 on implementation of a national curriculum in England and Wales stated:

> "Attainment targets will be set for all three core subjects of Mathematics, English and Science. These will establish what children should normally be expected to know, understand and be able to do at around the ages of 7, 11, 14 and 16, and will enable the progress of each child to be measured against established national standards. .... Targets must be sufficiently specific for pupils, teachers, parents, and others to have a clear idea of what is expected, and to provide a sound basis for assessment" [8].

It will be seen immediately that there is a potential contradiction in this specification for attainment targets. To be "sufficiently specific" they must be expressed using such words as "can do", "can calculate", "knows" and so can test only items of knowledge or skills. It is far from clear how a specific attainment target could be used to measure understanding in the above sense.

Just as it is not enough to test a list of items of knowledge, or a list of skills, so it is not enough to make a list of links between different mathematical concepts. While the existence of links is evidence of mathematical understanding, the absence of particular links cannot be used as evidence of lack of understanding. Since mathematical understanding is personal to the individual it is not possible to forecast in advance which particular links will be formed by each individual.

It follows that any attempt to assess understanding must concentrate instead on global demarcation of the area of stable knowledge and on the ability to move comfortably within this area. This implies tasks which encourage such movement, visiting not only those items of knowledge which are stable but also the boundaries at which knowledge becomes unstable or links have not yet been established. It is not at all clear how such tasks might be constructed, although investigations and projects give some indications, but it is certain that crude and specific "attainment targets" will not provide a suitable mechanism.

If, however, such assessment tasks are viewed not in terms of external examination but as a kind of self-assessment, through which students explore the extent of their area of stable knowledge, then they begin to define the same kind of activity as should enable the area of stable knowledge to increase. Thus assessment of understanding depends upon the same processes which encourage understanding to develop.

4. Encouragement of understanding

Where there is fragmentary or unstable knoweledge, how can more conceptual links be formed and the items of knowledge become more stable? How can the ability to move around an area of stable knowledge be developed? There is in fact very little research evidence on these points. Comparisons of mathematical education in different countries suggest that there are two extreme views. The first (practised, for example, in Taiwan) is to learn items of knowledge by rote, to have links between one item and another pointed out, to memorise all these facts, and then to practise a huge number of examples in which context or intrinsic parameters gradually become less similar to those of the original teaching. The second (as attempted in some UK schemes) is to begin with practical work through which concepts and links are gradually discovered, and then to have rules or procedures explained in terms of the conceptual constructs already established.

The difference between these two extremes may explain some of the differences in performance when the UK-oriented CSMS tests of Hart [10] were given to Taiwanese children [16]. Ratio problems involving small whole numbers were frequently reduced to an additive problem by the UK pupils but required accurate memory of a rule by the Taiwanese pupils. While the whole numbers remained small, the UK pupils had an advantage; when the ratios involved more complicated numbers the Taiwanese pupils who had mastered the rule retained their facility with the new problem [15].

In practise, no country taking one of the extreme positions outlined above has remained satisfied with it for very long. In the United Kingdom the tendency to precede the teaching of number facts by a large amount of practical work and guided discovery has been modified by the introduction of rules and memorisation at an earlier stage. The government announcement on "attainment targets" quoted above may be viewed as part of an extreme position (or at least was perceived to do so - it is doubtful whether it ever did). Conversely, in Taiwan there is an attempt to emphasize "teaching for understanding" as a reaction against teaching which emphasized too much rote learning at too early a stage. These reactions suggest that the optimal system is a mixture of taught rules and personal discovery.

The description of mathematical understanding as a sufficiently large connected area of stable knowledge, together with the ability to traverse this area mentally making connections and selecting procedures, does however suggest that taught rules and practical activities are not sufficient. The teaching must include features which actively encourage movement around the area, testing stability of knowledge and forming conceptions. The elements of mathematics teaching recommended by the Cockcroft Committee - in particular discussion between teacher and pupils and between pupils themselves, appropriate practical work, problem-solving including applications to everyday situations, and investigational work - may be viewed as strategies towards this end. So may the use of illuminating diagrams, the improvement of notation, the introduction of new symbolism, or the deliberate leaving of loose ends in a topic. The Cockcroft Report, perhaps unintentionally, gives primacy to discussion:

> "The ability to "say what you mean and mean what you say" should be one of the outcomes of good mathematics teaching. This ability develops as a result of opportunities to talk about mathematics, to explain and discuss results which have been obtained, and to test hypotheses. Moreover, the many different topics which exist within mathematics at both primary and secondary level should be presented and developed in such a way that they are seen to be inter-related. Pupils need the explicit help, which can only be given by extended discussion, to establish these relationships; even pupils whose mathematical attainment is high do not easily do this for themselves" [4] §246.

The claim that *only* extended discussion can achieve these results is absurd. But, even if the Cockcroft Committee intended to say that discussion is *one* way of establishing relationships, the research evidence is not clear. Empirical work suggests that all the teaching strategies mentioned above, including discussion, may have the desired effects for some pupils but may confuse other pupils [18]. Much more research is needed before claims are made about teaching strategies which encourage the development of understanding for all pupils.

Where such strategies as discussion, investigational work, practical work or problem solving have been successful in encouraging the development of understanding there may be an underlying common factor: teaching which arouses the student's personal enthusiasm is more likely to encorage the development of understanding. Clearly it is not enough to generate personal enthusiasm for watching the teacher perform. The enthusiasm must be of a kind that leads to personal exploration - first perhaps with practical materials and then mentally. Exploration of one's own knowledge and of possible conceptual links would seem to be the key to development of understanding. I suspect this is not a function of any particular teaching style or strategy - for example, the holding of class discussions might both encourage and inhibit such exploration - but that it is a function of the degree of enthusiasm for the task generated by the teaching. Which is, to return to the context of this collection of papers, simply to articulate what I realised when I first attended lectures by Christopher Zeeman thirty years ago.

References

[1] Bergeron, A., Herscovics, N., & Bergeron, J.C. "Kindergartners' knowledge of numbers: a longitudinal case study. Part II: Abstraction and formalisation" *Proceedings of the 11th International Conference on Psychology of Mathematics Education*, vol. II, p.352-360, Montreal 1987.

[2] Buxton, L. "Four levels of understanding" *Mathematics in Schools 7* (4), (September 1978), 36.

[3] Byers, V. & Herscovics, N. "Understanding school mathematics" *Mathematics Teaching* 81 (December 1977), 24-27.

[4] Cockcroft, W.H. *Mathematics counts.* Report of the Committee of Inquiry into the Teaching of Mathematics in Schools, HMSO, London 1982.

[5] Denvir, B. *The development of number concepts in low-attainers in mathematics aged 7-9 years*, Ph.D. thesis, University of London.

[6] Denvir, B. & Brown, M. "The feasibility of class-administered diagnostic assessment in primary mathematics" *Education Research* 29 (1987), 95-107.

[7] Denvir, B., Brown, M. & Eve, P. *Attainment targets and Assessment in the Primary Phase: Mathematics Feasibility Study*, Department of Education and Science, London 1987.

[8] Department of Education and Science. *The National Curriculum 5-16: A Consultation Document.* London 1987.

[9] Haylock, D.W. "Understanding Mathematics: Making Connections" *Mathematics Teaching* 98 (March 1982) 54-56.

[10] Hart, K.M. *Children's Understanding of Mathematics: 11-16*, John Murray, London 1981.

[11] Herscovics, N. & Bergeron, J.C. "Psychological questions regarding a new model of understanding elementary school mathematics" *Proceedings of 3rd Annual Meeting of North American Chapter of International Group for the Psychology of Mathematics Education*, p.69-76, 1981.

[12] Herscovics, N. & Bergeron, J.C. "A constructivist model of understanding" *Proceedings of 4th Annual Meeting of North American Chapter of International Group for the Psychology of Mathematics Education*, p.28-35, 1982.

[13] Herscovics, N., Bergeron, J.C. & Bergeron, A., "Kindergartners' knowledge of numbers: a longitudinal case study. Part I: Intuitive and procedural understanding", *Proceedings of 11th International Conference on Psychology of Mathematics Education* Vol. II, p.344-351, Montreal 1987.

[14] Hershkowitz, R. (to appear).

[15] Lin, F.L. (to appear).

[16] Lin, F.L., Kuo, F.P. & Lin, K.H. *Junior High School Children's Understanding of Ratio and Proportion.* Department of Mathematics, Taiwan Normal University (in Chinese), 1985.

[17] Pirie, S.E.B. "Understanding: instrumental, relational, intuitive, constructed, formalised ...? How can we know?" *For the learning of mathematics*, 8 (1988).

[18] Pirie, S.E.B. & Schwarzenberger, R.L.E. "Mathematical discussion and mathematical understanding" *Educational Studies in Mathematics*, 19 (1988).

[19] Schroeder, T.L. "Students' understanding of mathematics: a review and synthesis of some recent research" *Proceedings of 11th International Conference on Psychology of Mathematics Education* Vol. III, p.332-338, Montreal 1987.

[20] Schwarzenberger, R.L.E. *Targets for Mathematics in Primary Education*, Warwick Seminar on Public Education Policy. Trentham Books, Stoke-on-Trent 1987.

[21] Skemp, R.R. "Relational and instrumental understanding" *Mathematics Teaching* 77 (December 1976), 20-26.

[22] Skemp, R.R. "Goals of learning and qualities of understanding" *Mathematics Teaching* 88 (September 1979), 44-49.

[23] Skemp, R.R. *Intelligence, Learning and Actions: A Foundation of Theory and Practise in Education.* John Wiley, Chichester 1979.

[24] Skemp, R.R. "Symbolic understanding" *Mathematics Teaching* 99 (June 1982), 59-61.

[25] Tall, D. "The dynamics of understanding mathematics" *Mathematics Teaching* 84 (September 1978), 50-52.

[26] Vinner, S. "True instrumental versus quasi-instrumental performance in mathematics learning" *Papers in Honour of Richard Skemp*, Mathematics Education Research Centre, University of Warwick, 1986.

And God Said, "$0 \notin \mathbb{N}$"

Researched by Tim Auckland

I have found the answer to the longstanding question "Is zero a natural number?" by accident (?) while reading the Book of Ecclesiastes.

מְעֻוָּת לֹא־יוּכַל לִתְקֹן וְחֶסְרוֹן
לֹא־יוּכַל לְהִמָּנוֹת׃

15 *That which is* crooked cannot be made straight: and that which is wanting cannot be numbered.

Ecclesiastes 1 v.15, second part, clearly states '$0 \notin \mathbb{N}$'.

It has been suggested that the translation is incorrect due to the translator not being a mathematician. Hence I have included the Hebrew version so that the reader can plainly see there is no error.

The alternative proposed was

'That which cannot be numbered is wanting'

or

'$\nexists$ uncountable sets.'

A second argument points to Matthew 5 v.17:

17 *Μη νομισητε ὁτι ἠλθον καταλυσαι τον νομον ἠ τους προφητας οὐκ ἠλθον καταλυσαι ἀλλα πληρωσαι.*

'Do not suppose that I have come to abolish the Law and the prophets; I did not come to abolish, but to complete.'

With the suggestion that $0 \in \mathbb{N}$ is the completion of the Law. However Jesus himself refutes this in the next verse:

18 *αμην γαρ λεγω ὑμιν, ἑως ἀν παρελθη ὁ οὐρανος και ἡ γη, ἰωτα ἑν ἡ μια κεραια οὐ μη παρελθη ἀπο του νομου ἑως ἀν παντα γενηται.*

'I tell you this: so long as heaven and earth endure, not a letter, not a stroke, will disappear from the Law until all that must happen has happened.'

Thus, as it states in Genesis, the world began on the *first* day. We shall have to wait and see if it ends on the ω–th day.

A Lattice of Topologies

Simon Morris

Let $\mathcal{T}_1$, $\mathcal{T}_2$ be topologies on a set X, so

$$\mathcal{T}_1, \mathcal{T}_2 \subseteq \mathsf{P}X.$$

Then $\mathcal{B} = \{\mathcal{O}_1 \cap \mathcal{O}_2 \mid \mathcal{O}_1 \in \mathcal{T}_1, \mathcal{O}_2 \in \mathcal{T}_2\}$ is the base for a topology $\mathcal{T}_1 \vee \mathcal{T}_2$ on X, as $X \in \mathcal{B}$ (so $\mathcal{B}$ covers X), and

$$\begin{aligned}\mathcal{O}, \mathcal{O}' \in \mathcal{B} &\Rightarrow \mathcal{O} = \mathcal{O}_1 \cap \mathcal{O}_2, \mathcal{O}' = \mathcal{O}_1' \cap \mathcal{O}_2' \\ &\Rightarrow \mathcal{O} \cap \mathcal{O}' = (\mathcal{O}_1 \cap \mathcal{O}_1') \cap (\mathcal{O}_2 \cap \mathcal{O}_2') \in \mathcal{B}.\end{aligned}$$

$\mathcal{T}_1 \vee \mathcal{T}_2$ is the topology on X generated by the subbase $\mathcal{T}_1 \cup \mathcal{T}_2$. $\mathcal{T}_1 \wedge \mathcal{T}_2 = \mathcal{T}_1 \cap \mathcal{T}_2$ is trivially a topology.

Let $\mathcal{T}_X$ be the set of topologies on X,

$$\mathcal{T}_X^0 = \{\emptyset, X\}$$

the indiscete topology on X, and

$$\mathcal{T}_X^1 = \mathsf{P}X$$

the discrete topology on X.

Proposition *$\mathcal{T}_X$ is a lattice under the partial order of set inclusion, infimum $\wedge$, and supremum $\vee$. It has lower bound $\mathcal{T}_X^0$ and upper bound $\mathcal{T}_X^1$.*

Proof Pick $\mathcal{T}_1, \mathcal{T}_2 \in \mathcal{T}_X$. Then

$$\mathcal{T}_1, \mathcal{T}_2 \geq \mathcal{T}_1 \wedge \mathcal{T}_2 = \mathcal{T}_1 \cap \mathcal{T}_2,$$

and

$$\mathcal{T}_1, \mathcal{T}_2 \geq \mathcal{T}_3 \Rightarrow \mathcal{T}_1 \wedge \mathcal{T}_2 = \mathcal{T}_1 \cap \mathcal{T}_2 \geq \mathcal{T}_3,$$

$\therefore \wedge$ is an infimum.

$$\mathcal{T}_1, \mathcal{T}_2 \leq \mathcal{T}_1 \vee \mathcal{T}_2$$

as any open set in $\mathcal{T}_i$ is in the base generating $\mathcal{T}_1 \vee \mathcal{T}_2$;

$$\mathcal{T}_1, \mathcal{T}_2 \leq \mathcal{T}_3 \Rightarrow \mathcal{T}_1 \vee \mathcal{T}_2 \leq \mathcal{T}_3$$

as every element of the base generating $\mathcal{T}_1 \vee \mathcal{T}_2$ is open in $\mathcal{T}_3$. $\therefore \vee$ is a supremum.

$$\forall \mathcal{T} \in \mathcal{T}_X, \quad \mathcal{T}_X^0 \leq \mathcal{T} \leq \mathcal{T}_X^1$$

is trivial. $\square$

A lattice L is said to be complete if every non-empty subset of L has a supremum and an infimum. There is a standard result:

Proposition *If L is a lattice with upper bound 1 such that every non-empty subset of L has an infimum, then L is complete.*
Proof Pick $\mathcal{S} \subseteq L$; we construct a supremum for $\mathcal{S}$.
Let $\mathcal{U}$ be the set of upper bounds of $\mathcal{S}$. $1 \in \mathcal{U}$, so $\mathcal{U} \neq \emptyset$; set $u = \inf \mathcal{U}$.
Every $s \in \mathcal{S}$ is a lower bound for $\mathcal{U}$, so $u \geq s$ $(\forall s \in \mathcal{S})$, so u is an upper bound for S.
If also u' is an upper bound for $\mathcal{S}$, then $u' \in \mathcal{U}$, so $u \leq u'$.
$\therefore$ u is a supremum for $\mathcal{S}$. □

Proposition *$\mathcal{T}_X$ is complete.*
Proof ETS every non-empty subset of L has an infimum.
Pick $\mathcal{S} \subseteq \mathcal{T}_X, \mathcal{S} \neq \emptyset$. Set $\mathcal{T} = \cap\mathcal{S}$.
It is easy to see that $\mathcal{T}$ is a topology on X, so $\mathcal{T} \in \mathcal{T}_X$, and $\mathcal{T}$ is certainly a lower bound for $\mathcal{S}$.
If $\mathcal{T}'$ is a lower bound for $\mathcal{S}$, then

$$\mathcal{T}' \leq \cap\mathcal{S} = \mathcal{T}.$$

$\therefore$ $\mathcal{T}$ is an infimum for S. □

Chess Puzzle

Adam Chalcraft

How many of the 32 chess pieces in a normal set can be placed on a chess board so that no piece attacks or defends any other?

(The position they are placed in need not be legally reachable - but remember the direction of play when placing pawns!)

A, B, C, D Problem

Luke Reader

(i) Find four digits (i.e. integers between 0 and 9 inclusive) such that $ABCD$ (i.e. $1000 \times A + 100 \times B + 10 \times C + D$) is equal to $A^B \cdot C^D$.

(ii) Prove that the solution to (i) is unique, without using number-crunching.

On Phone Numbers and Diophantine Approximation

Christopher J. Fewster

Question [1]: is there a positive integer n such that 2^n, written in base 10, begins with a 7?

Answer: yes, in fact there are infinitely many such n, but one has to look surprisingly far to find the first example ($n = 46$).

This article will concern itself with such profound mathematical thoughts as "Is there a prime number whose most significant digits are the same as my phone number?" (there is) and tries to avoid potentially embarrassing questions such as "If such a prime number exists, how big is it?" (very big, probably). In particular, we will examine sets and sequences which have the property that *any* finite sequence of digits may be found as the first few most significant digits of one of their members. (So even E.T. can rest assured that his phone number, with an n-digit intergalactic dialling code, is still represented as the opening digits of a prime.)

This property is equivalent to asserting that the fractional parts of the logarithm of members of the set or sequence are dense in $[0,1)$, where we define the fractional part of x: $\text{frac}\ x = x - \text{int}\ x$ and $\text{int}\ x$ is the unique integer such that $x - 1 < \text{int}\ x \leq x$ †, so that, for example, frac -0.75 = 0.25. The equivalence becomes clear when we realise that for any $x \in \mathbf{R}^+$ and $r \geq 1$, there exists ϵ_r such that $y \in \mathbf{R}^+$ has the same first r opening digits as x if and only if $\text{frac}\ \log_{10} y \in [\text{frac}\ \log_{10} x, \text{frac}\ \log_{10} x + \epsilon_r)$. The above motivates the following definition.

Definition 1 A set or sequence $A \subseteq \mathbf{R}^+$ is logarithmically dense to base b or $\log_b$-dense if and only if $\{\text{frac}\ \log_b x \mid x \in A\}$ is dense in $[0,1)$.

In this article, we will use the case $b = 10$ as our main example, though the theory is quite general. Note that we can extend Definition 1 to higher dimensions by defining $\text{frac}\ \log_b(x_1, x_2, \ldots, x_n) = (\text{frac}\ \log_b x_1, \text{frac}\ \log_b x_2, \ldots, \text{frac}\ \log_b x_n)$ and requiring $\{\text{frac}\ \log_b \underline{x} \mid \underline{x} \in A\}$ to be dense in $[0,1)^n$ where $A \subseteq (\mathbf{R}^+)^n$.

It will be useful to bear in mind that we can use $[0,1)$ to coordinatize the circle S^1 and therefore a sequence $\{x_n\}$ in $\mathbf{R}^+$ induces an orbit of the circle under the transformation $x \rightarrow \text{frac}\ \log_b x$. We now introduce a useful concept which will pop up in most of the proofs ahead.

Definition 2 Let A be a set or sequence in $\mathbf{R}^+$. Then define $\text{Mesh}_b A$ by

$$\text{Mesh}_b A = \sup_{x \in [0,1)} \inf_{y \in A} \text{d}(x, \text{frac}\ \log_b y)$$

where

$$\text{d}(p,q) = \begin{cases} q - p & \text{if } q \geq p, \\ 1 + q - p & \text{if } q < p \end{cases}$$

† I use the non-standard notation frac and int rather than the usual { } and [ ] to prevent confusion with set notation. I have also restrained myself, despite severe temptation, from writing frac log x as flog x. One notational abuse per article is quite enough.

so that for any $p, q \in [0,1)$, $\mathrm{d}(p,q) \geq 0$ and hence $\mathrm{Mesh}_b A \geq 0$. (If we think about the association of [0,1) with S^1 then $\mathrm{d}(p,q)$ is the "angle" between p and q, measured in the positive sense, where our angular measure runs from 0 to 1 instead of from 0 to 2π. For example, $\mathrm{d}(0.1, 0.2) = 0.1$, but $\mathrm{d}(0.2, 0.1) = 0.9$.)

It is clear that A is $\log_b$-dense if and only if $\mathrm{Mesh}_b A = 0$, and also that if $B \subseteq A$, then $\mathrm{Mesh}_b B \geq \mathrm{Mesh}_b A$. Further, if the finite sequence $x_1, x_2, ..., x_n$ corresponds to at least one complete circuit of S^1 (not necessarily returning to the starting point) in positive steps of less than ϵ, i.e. $\mathrm{d}(\text{frac}\log_b x_r, \text{frac}\log_b x_{r+1}) < \epsilon$, then we have $\mathrm{Mesh}_b\{x_1, x_2, ..., x_n\} < \epsilon$ (and the same applies if the steps are in the negative sense). Quite a few of the proofs in this article that a sequence $\{x_n\}$ is $\log_b$-dense hinge on showing that for any $\epsilon > 0$, there is a subsequence $\{x_{n_r}\}$ such that $\mathrm{Mesh}_b\{x_{n_r}\} < \epsilon$. We have $\mathrm{Mesh}_b\{x_n\} \leq \mathrm{Mesh}_b\{x_{n_r}\} < \epsilon$, so since ϵ is arbitrary, $\mathrm{Mesh}_b\{x_n\} = 0$ and so $\{x_n\}$ is $\log_b$-dense.

Having dispensed with the preliminaries, we are ready to prove some results. First, we meet a set of $\log_b$-dense sequences.

Theorem 1 *If $k > 0$ and $\log_b k \notin \mathbb{Q}$ then $\{k^n \mid n \in \mathbb{N}\}$ is $\log_b$-dense.*

Proof From elementary Diophantine approximation theory [2], since $\log_b k \notin \mathbb{Q}$, we know that there are infinitely many rationals p/q such that $|\log_b k - p/q| < 1/q^2$. It is clear that we can find such p/q with q arbitrarily large, so for any $\epsilon > 0$ we have $q \in \mathbb{N}, p \in \mathbb{Z}$ such that $q > 1/\epsilon$ and $|\log_b k - p/q| < 1/q^2$. Thus $|q\log_b k - p| < 1/q < \epsilon$. This means that the sequence $\{k^{qr} \mid r \in \mathbb{N}\}$ represents an orbit of the circle in steps (of constant sign) with magnitude less than ϵ, and we see that $\mathrm{Mesh}_b\{k^{qr}\} < \epsilon$. (For this orbit will complete an entire circuit in a finite number of steps and we use our earlier observation). But $\mathrm{Mesh}_b\{k^r\} \leq \mathrm{Mesh}_b\{k^{qr}\} < \epsilon$ and ϵ was arbitrary, so $\mathrm{Mesh}_b\{k^r\} = 0$. Hence $\{k^r\}$ is $\log_b$-dense. $\square$

For example, we have proved that the powers of 2 form a $\log_{10}$-dense sequence (and thus that there is a power of 2 beginning with a 7) because $\log_{10} 2$ is irrational (otherwise $2 = 10^{p/q}$, some $p, q \geq 1$, so $2^q = 10^p = 5 \times 2 \times 10^{p-1}$, and 2^q is therefore divisible by 5, which would *never* do). The other point to make is that I could have proved Theorem 1 by observing that $\{k^n \mid n \in \mathbb{N}\}$ for $\log_b k \notin \mathbb{Q}$ induces an irrational orbit of the circle, and that such orbits are dense, but I wanted to justify the presence of "Diophantine approximation" in the title.

We now move on to a lemma, which, although quite simple, is fundamental to the remainder of this article. On a point of notation, I use $y_n \to y^+$ to mean $y_n \to y$ and $y_n > y$ for all n (note that the inequality is strict). $y_n \to y^-$ has a corresponding definition.

Lemma 1 *The following hold:*
(a) If $\log_b x_{n+1} - \log_b x_n \to \delta$ where $0 < \delta < 1$ then $\mathrm{Mesh}_b\{x_n\} \leq \delta$;
(b) If $\log_b x_{n+1} - \log_b x_n \to 0^+$ and $x_n \to \infty$ then $\mathrm{Mesh}_b\{x_n\} = 0$;
(c) If $\log_b x_{n+1} - \log_b x_n \to 0^-$ and $x_n \to 0$ then $\mathrm{Mesh}_b\{x_n\} = 0$.

Proof (a) For any $\epsilon > \delta$, there exists n_0 such that for $n \geq n_0$, $\delta/2 < \log_b x_{n+1} - \log_b x_n < \epsilon$. We consider $x_{n_0}, ..., x_{n_0+1+\text{int}(2/\delta)}$, and the induced orbit on S^1. The step-length is between $\delta/2$ and ϵ and so in $1 + \text{int}(2/\delta)$ steps we complete at least one entire circuit of the circle in steps of less than ϵ. Hence $\mathrm{Mesh}_b\{x_n\} \leq \mathrm{Mesh}_b\{x_{n_0}, ..., x_{n_0+1+\text{int}(2/\delta)}\} < \epsilon$. So $\mathrm{Mesh}_b\{x_n\} < \epsilon$ for any $\epsilon > \delta$, so $\mathrm{Mesh}_b\{x_n\} \leq \delta$.

(b) For any $\epsilon > 0$, note that we have n_0 such that for $n \geq n_0$, $0 < \log_b x_{n+1} - \log_b x_{n_0} < \epsilon$. Since $x_n \to \infty$, there exists r (greater than $1/\epsilon$, without loss of generality) such that $\log_b x_{n_0+r} - \log_b x_{n_0} > 1$. Therefore, the $r+1$ points $x_{n_0}, \ldots, x_{n_0+r}$ correspond to at least one circuit of S^1 with a step-length of less than ϵ, so $\mathrm{Mesh}_b\{x_n\} \leq \mathrm{Mesh}_b\{x_{n_0}, \ldots, x_{n_0+r}\} < \epsilon$. Hence we have $\mathrm{Mesh}_b\{x_n\} < \epsilon$ for any $\epsilon > 0$, so $\mathrm{Mesh}_b\{x_n\} = 0$.

(c) Proceed as in (b) except rotation takes place in the opposite sense. The condition $x_n \to 0$ is necessary to achieve $\log_b x_{n_0+r} - \log_b x_{n_0} < -1$ at the appropriate point. □

Corollary *The following hold:*
(a) If $\log_b x_{n+1} - \log_b x_n \to c$ *and* $\mathrm{frac}\ c \neq 0$ *then* $\mathrm{Mesh}_b\{x_n\} \leq \mathrm{frac}\ c$;
(b) If $\log_b x_{n+1} - \log_b x_n \to c^+$, $c \in \mathbf{Z}$ *and* $x_n/b^{cn} \to \infty$ *then* $\mathrm{Mesh}_b\{x_n\} = 0$;
(c) If $\log_b x_{n+1} - \log_b x_n \to c^-$, $c \in \mathbf{Z}$ *and* $x_n/b^{cn} \to 0$ *then* $\mathrm{Mesh}_b\{x_n\} = 0$.
Proof In each case put $y_n = x_n/b^{n \cdot \mathrm{int}(c)}$. Then:
(a) $\log_b y_{n+1} - \log_b y_n \to \mathrm{frac}\ c \neq 0 \quad 0 < \mathrm{frac}\ c < 1$. Apply Lemma 1(a).
(b) $\log_b y_{n+1} - \log_b y_n \to 0^+$ and $y_n \to \infty$. Apply Lemma 1(b).
(c) $\log_b y_{n+1} - \log_b y_n \to 0^-$ and $y_n \to 0$. Apply Lemma 1(c).
In this way we obtain $\mathrm{Mesh}_b\{y_n\}$. But $\mathrm{Mesh}_b\{y_n\} = \mathrm{Mesh}_b\{x_n\}$ as Mesh_b depends only on the fractional part of the logarithm and $\mathrm{frac}\ \log_b x_n = \mathrm{frac}\ \log_b y_n$. Thus in each of the cases (a), (b), (c) we get the required result. □

Lemma 1 and its corollary allow us to generalize Theorem 1 considerably. But first we prove the result about prime numbers you've all been waiting for...

Theorem 2 *The sequence of prime numbers* $\{p_n\}$ *is logarithmically dense to any base.*
Proof It is a well-known result that $p_{n+1}/p_n \to 1^+$, following immediately from the Prime Number Theorem (see [3]). Therefore $\log_b p_{n+1} - \log_b p_n \to 0^+$. It is an even better known result that there are infinitely many primes, so that $p_n \to \infty$. Thus by Lemma 1(b) we have $\mathrm{Mesh}_b\{p_n\} = 0$, so $\{p_n\}$ is $\log_b$-dense, and b was arbitrary. □

Clearly, by Lemma 1(b) any increasing sequence $\{x_n\}$ with $x_{n+1}/x_n \to 1$ and $x_n \to \infty$ is logarithmically dense to any base. This includes very many sequences: all sufficiently slowly increasing sequences which tend to infinity. For example, the sequences of squares, cubes and in general r-th powers come into this category as $(n+1)^r/n^r = (1+1/n)^r \to 1^+$ as $n \to \infty$. The next item on the menu is a generalisation of Theorem 1 which deals with exponentially growing sequences.

Theorem 3 *If* $x_{n+1}/x_n \to k > 0$ *and* $\log_b k \notin \mathbf{Q}$ *then* $\{x_n\}$ *is* $\log_b$*-dense.*
Proof We have $x_{(n+1)r}/x_{nr} \to k^r$ as $n \to \infty$ and so $\log_b x_{(n+1)r} - \log_b x_{nr} \to \log_b k^r$, for any $r \in \mathbf{N}$. Now we have $\log_b k \notin \mathbf{Q}$ so $\mathrm{frac}\ \log_b k^r = \mathrm{frac}\ r\log_b k \neq 0$ for all r. Thus by corollary (a) to Lemma 1, $\mathrm{Mesh}_b\{x_{nr} \mid n \in \mathbf{N}\} \leq \mathrm{frac}\ \log_b k^r$ for all r. So $\mathrm{Mesh}_b\{x_n\} \leq \mathrm{frac}\ \log_b k^r$ for all r. But by Theorem 1, $\{k^r\}$ is $\log_b$-dense as $\log_b k \notin \mathbf{Q}$. So for any $\epsilon > 0$ there exists r such that $\mathrm{frac}\ \log_b k^r < \epsilon$. Hence $\mathrm{Mesh}_b\{x_n\} < \epsilon$ for any $\epsilon > 0$, and we obtain $\mathrm{Mesh}_b\{x_n\} = 0$. □

All is not, however, lost for sequences $\{x_n\}$ such that $x_{n+1}/x_n \to k$ where $\log_b k \in \mathbf{Q}$.

Theorem 4 *(a) If* $x_{n+1}/x_n \to k^+$, $\log_b k \in \mathbf{Q}$ *and* $x_n/k^n \to \infty$ *then* $\{x_n\}$ *is* $\log_b$*-dense;*
(b) If $x_{n+1}/x_n \to k^-$, $\log_b k \in \mathbf{Q}$ *and* $x_n/k^n \to 0$ *then* $\{x_n\}$ *is* $\log_b$*-dense.*

Proof We have $\log_b k = p/q$, and so we consider the subsequence $\{x_{nq}\}$. $\log_b x_{(n+1)q} - \log_b x_{nq} \to p^{\pm}$ and $x_{nq}/k^{nq} \to {\infty \atop 0}$, i.e. $x_{nq}/b^{pn} \to {\infty \atop 0}$, so by corollary (b) or (c) to Lemma 1, $\mathrm{Mesh}_b\{x_n\} \leq \mathrm{Mesh}_b\{x_{nq}\} = 0$. □

Theorem 3 and the more restrictive Theorem 4 encompass a wide range of exponentially growing sequences: for example $x_n = n^r k^n$ for any $n \in \mathbb{N}$ and almost all k (all k if $r \neq 0$). We can also see that the Fibonacci sequence f_n is $\log_b$-dense to any integer base b (we define $f_0 = 0, f_1 = 1$ and $f_{n+2} = f_{n+1} + f_n$). For, by solving the recurrence relation, we obtain

$$f_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right) \text{ so } \frac{f_{n+1}}{f_n} \to \frac{1+\sqrt{5}}{2}.$$

Now suppose there were an integer $b > 0$ such that $\log_b \frac{1+\sqrt{5}}{2} \in \mathbb{Q}$; then we would have $b^p = (\frac{1+\sqrt{5}}{2})^q$ with $p, q \in \mathbb{N}, q \neq 0$. However b^p is an integer, whereas we can verify quite easily that $(\frac{1+\sqrt{5}}{2})^q = \alpha + \beta\sqrt{5}$ with $\alpha, \beta \in \mathbb{Q}, \beta \neq 0$, for all $q > 0$, and so $(\frac{1+\sqrt{5}}{2})^q$ is not an integer. This contradiction allows us to infer that $\log_b(\frac{1+\sqrt{5}}{2}) \notin \mathbb{Q}$ for all integer bases b, so by Theorem 3, we see that $\{f_n\}$ is logarithmically dense in all positive integer bases.

It is clear that in a similar way, many sequences which satisfy linear recurrence relations are logarithmically dense.

We now move on to what I think is quite a surprising result, concerning a certain subset of log-dense sequences. It draws upon the concept of "uniformly distributed sequences" [4], the theory of which was first promulgated by H. Weyl and others in the 1920s. The motivation behind that theory was to go beyond the concept of a densely distributed sequence in [0,1) to a sequence in [0,1) whose points are in some sense evenly or uniformly distributed.

Weyl defined a sequence ω_n on [0,1) to be uniformly distributed if for all intervals $J \subseteq [0,1)$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \chi_J(\omega_n) = \ell(J)$$

where $\ell(J)$ is the length of J and χ_J is its characteristic function,

$$\chi_J(x) = \begin{cases} 1 & \text{if } x \in J \\ 0 & \text{if } x \notin J. \end{cases}$$

The first important result in the field was the Weyl Criterion, which can be used to determine whether a given sequence is uniformly distributed. This is stated as: ω_n is uniformly distributed if and only if $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i h\omega_n} = 0$ for any non-zero integer h. Hlawka [4] gives a proof of this result (pp. 4-9) - the proof is quite simple and the result gains its depth from an application of the Weierstrass Approximation Theorem. We now modify the definition and the Weyl criterion to suit our purposes.

Definition 3 $\{x_n\} \subseteq \mathbb{R}^+$ is uniformly logarithmically dense to base b if and only if for any $[x, x+\epsilon) \subseteq [0,1)$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \chi_{[x,x+\epsilon)}(\mathrm{frac}\ \log_b x_n) = \epsilon.$$

It is clear that if $\{x_n\}$ is uniformly $\log_b$-dense , then $\{x_n\}$ is $\log_b$-dense . The Weyl criterion becomes:

Theorem 6 *$\{x_n\}$ is uniformly $\log_b$-dense if and only if*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h \log_b x_n} = 0$$

for any non-zero integer h. □

Note that the absence of the frac function does not matter here, as $e^{2\pi i h k} = 1$ for any $k \in \mathbf{Z}$, and so $e^{2\pi i h \text{frac } x} = e^{2\pi i h x}$.

To show that this concept is not empty, we use Theorem 6 to re-prove Theorem 1, and indeed to strengthen it.

Theorem 7 *If $k > 0$ and $\log_b k \notin \mathbf{Q}$ then $\{k^n\}$ is uniformly $\log_b$-dense .*

Proof Observe that

$$\sum_{n=1}^{N} e^{2\pi i h \log_b x_n} = \sum_{n=1}^{N} e^{2\pi i h n \log_b k} = \sum_{n=1}^{N} (e^{2\pi i h \log_b k})^n$$
$$= \frac{e^{2\pi i h \log_b k}(1 - (e^{2\pi i h \log_b k})^N)}{1 - e^{2\pi i h \log_b k}}.$$

Now $\log_b k \notin \mathbf{Q}$ so $e^{2\pi i h \log_b x_n} \neq 1$ for all $h \in \mathbf{Z} \setminus \{0\}$ and so the sum converges. Therefore we can write $\sum_{n=1}^{N} e^{2\pi i h \log_b x_n} = M(1 - e^{2\pi i h N \log_b k})$ for finite $|M(h,k)|$. Hence $|\sum_{n=1}^{N} e^{2\pi i h \log_b x_n}| < 2M$ for all N, and so $\lim_{N\to\infty} 1/N \sum_{n=1}^{N} e^{2\pi i h \log_b x_n} = 0$ for all non-zero integers h, and the Weyl criterion is satisfied. □

We can generalize the concept of uniform logarithmic density to higher dimensions and also get a generalized Weyl criterion.

Definition 4 $\{x_n\} \subseteq (\mathbf{R}^+)^m$ is uniformly $\log_b$-dense if and only if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi_J(\text{frac } \log_b \underline{x}_n) = v(J)$$

where we define frac $\log_b \underline{x}_n$ as earlier, J is a subset of $[0,1)^m$ of the form $J = [\alpha_1, \alpha_1 + \epsilon_1) \times [\alpha_2, \alpha_2 + \epsilon_2) \times \ldots \times [\alpha_m, \alpha_m + \epsilon_m)$ and $v(J) = \epsilon_1 \epsilon_2 \ldots \epsilon_m$ is the "volume" of J. χ_J is the characteristic function of J.

Theorem 8 (Weyl criterion) *$\{\underline{x}_n\} \subseteq (\mathbf{R}^+)^m$ is uniformly $\log_b$-dense if and only if*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i \underline{h} \cdot \log_b \underline{x}_n} = 0$$

for any $\underline{h} \in \mathbf{Z}^m \setminus \{\underline{0}\}$.

Proof Similar to the one-dimensional case - see Hlawka [4]. □

If we have another look at the proof of Theorem 7, we see that the critical step was to note that $h\log_b x_n$ was never an integer. In the same way, if we consider the sequence $\underline{x}_n = (k_1^n, k_2^n, ..., k_m^n)$, with $(k_1, k_2, ..., k_m) = \underline{k}$ chosen so that $\underline{h}\cdot\log_b\underline{k} \notin \mathbf{Z}$ for any $\underline{h} \in \mathbf{Z}^m\backslash\{0\}$, we see that $\underline{h}\cdot\log_b\underline{x}_n = n\underline{h}\cdot\log_b\underline{k}$ and all of the argument in Theorem 7 carries through. Thus $\{\underline{x}_n\}$ is uniformly $\log_b$-dense . Specialising to the case $b = 10$, and setting $\underline{k} = (3, 5, 7, 11, ...)$ so that each component is a prime, we note that $\underline{h}\cdot\log_{10}\underline{k} \in \mathbf{Z}$ implies that $3^{h_1}\cdot 5^{h_2}\cdot 7^{h_3}... = 10^s$ for some $x \in \mathbf{Z}$ (and in fact $s \geq 0$ without loss of generality) which is impossible, by uniqueness of prime factorization. (For we have asserted that $10^s = p/q$, where p is formed from the factors with positive h_i and q from those with negative h_i. Thus $p = 10^s q$, but p and q have no common factors if $q \neq 1$, and if $q = 1, s \neq 0$ then p is not divisible by 2 and we again reach a contradiction. The only solution is $p = q = 1$, $s = 0$, but this implies $\underline{h} = 0$, a case we have excluded.) We have thus proved that given an m-tuple of positive reals $(n_1, ..., n_m)$, there exists an $n \in \mathbf{N}$ such that 3^n begins with the digits of n_1, 5^n begins with the digits of n_2, and so on. Less formally, we see that there exists an n so that 3^n starts with your phone number, 5^n with your National Insurance number, 7^n with your height to the nearest cm, 11^n with the largest known prime...

This is rather surprising. The next theorem is even more so.

Theorem 9 *Suppose that* $\{x_n\} \subseteq (\mathbf{R}^+)^m$ *is uniformly* $\log_b$*-dense. Then* $\{y_n\} \subseteq (\mathbf{R}^+)^{m+1}$ *is* $\log_b$*-dense, where* $\underline{y}_n = (x_{n_1}, x_{n_2}, ..., x_{n_m}, n)$.

Proof Since $\{x_n\}$ is uniformly $\log_b$-dense , we have by Definition 4 that for $J = [\alpha_1, \alpha_1 + \epsilon_1) \times ... \times [\alpha_m, \alpha_m + \epsilon_m) \subseteq [0, 1)^m$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \chi_J(\underline{\omega}_n) = \epsilon = \prod_{i=1}^{N} \epsilon_i$$

where ω_n = frac $\log_b\{x_n\}$. By elementary analysis, for any $\delta > 0$, there exists N_0 such that for $N > N_0$, we have

$$\epsilon - \delta < \frac{1}{N}\sum_{n=1}^{N} \chi_J(\underline{\omega}_n) < \epsilon + \delta.$$

Now let $\{n_r\}$ be the increasing sequence defined by $\sum_{n=1}^{n_r} \chi_J(\underline{\omega}_n) = r$ (i.e. $\chi_J(\underline{\omega}_n) = 1$ if and only if $n = n_r$ for some r). Then we have r_0 so that $n_r > N_0$ for $r > r_0$. Thus $\epsilon - \delta < \frac{r}{n_r} < \epsilon + \delta$ and $\epsilon - \delta < \frac{r}{n_{r+1}-1} < \epsilon + \delta$ for $r > r_0$. Hence

$$\frac{\epsilon - \delta}{\epsilon + \delta} = \frac{r}{\epsilon + \delta}/\frac{r}{\epsilon - \delta} < \frac{n_{r+1} - 1}{n_r} < \frac{r}{\epsilon - \delta}/\frac{r}{\epsilon + \delta} = \frac{\epsilon + \delta}{\epsilon - \delta}.$$

Now $\frac{\epsilon-\delta}{\epsilon+\delta} \to 1^-$ as $\delta \to 0$ and $\frac{\epsilon+\delta}{\epsilon-\delta} \to 1^+$ in the same limit. So for any $\eta > 0$, there exists δ such that $\max\{|\frac{\epsilon-\delta}{\epsilon+\delta} - 1|, |\frac{\epsilon+\delta}{\epsilon-\delta} - 1|\} < \eta$ and r_0 such that $r > r_0$ implies $\frac{\epsilon-\delta}{\epsilon+\delta} < \frac{n_{r+1}-1}{n_r} < \frac{\epsilon+\delta}{\epsilon-\delta}$, i.e. $|\frac{n_{r+1}-1}{n_r} - 1| < \eta$. So $\frac{n_{r+1}-1}{n_r} \to 1$ as $r \to \infty$ and, as $n_r \to \infty$ as $r \to \infty$, we have

$n_{r+1}/n_r \to 1$. But by definition, $n_{r+1} > n_r$, so $n_{r+1}/n_r \to 1^+$ and by Theorem 4(a) in the case $k = 1$, we see that $\{n_r\}$ is $\log_b$-dense . Thus, for any $[\alpha_{m+1}, \alpha_{m+1} + \epsilon_{m+1}) \subseteq [0,1)$ there is an r with frac $\log_b n_r \in [\alpha_{m+1}, \alpha_{m+1} + \epsilon_{m+1})$. By definition of the sequence $\{n_r\}$, $\underline{x}_{n_r} \in J$, so $\underline{y}_{n_r} \in J \times [\alpha_{m+1}, \alpha_{m+1} + \epsilon_{m+1})$. So we have shown that for any subset $J' = [\alpha_1, \alpha_1 + \epsilon_1) \times [\alpha_2, \alpha_2 + \epsilon_2) \times \ldots \times [\alpha_{m+1}, \alpha_{m+1} + \epsilon_{m+1})$ of $[0,1)^{m+1}$, there is an n such that $\underline{y}_n \in J'$, i.e. $\{\underline{y}_n\}$ is $\log_b$-dense . □

Corollary (a) *The natural numbers are not uniformly logarithmically dense to any base (although they are logarithmically dense to every base).*
(b) *If* $\{\underline{x}_n\} \subseteq (\mathbf{R}^+)^m$ *is uniformly* $\log_b$*-dense and* $\underline{y}_n$ *is formed from* $\underline{x}_n$ *as above, then* $\underline{y}_n$ *is not* uniformly $\log_b$*-dense.*
Proof (a) If $\{n\}$ were uniformly $\log_b$-dense , then by Theorem 9, $\{(n,n)\}$ would be $\log_b$-dense. However, this is clearly impossible as the points frac $\log_b(n,n)$ are all on the diagonal of the unit square $[0,1)\times[0,1)$ and so $\{\text{frac} \log_b(n,n)\}$ is not dense in $[0,1)\times[0,1)$.
(b) This follows by a similar argument on the basis that $\{(\underline{x}_n, n, n)\} \subseteq (\mathbf{R}^+)^{m+2}$ is not $\log_b$-dense. □

Note that Corollary (a) to Theorem 9 translates into a result in the theory of uniform distribution - namely that the fractional parts of $\log_b n$ are not uniformly distributed on $[0,1)$; and it is not immediately apparent how to verify this from the Weyl criterion. In general, Theorem 9 could be summed up in the well-worn phrase, "Buy n, get one free!", for not only can you find r such that $3^r, 5^r, 7^r, \ldots$ begin with those important numbers in your life, but r itself begins with yet another! However, this is beginning to sound like an advert in a colour supplement, so before I end up with a job in sales management I'd better draw things to a close.

As I mentioned earlier, this article has so far studiously avoided any specific examples of primes beginning with significant phone numbers, so I'll conclude with three examples to demonstrate the size of the numbers necessary to get reasonable approximations. The examples are all based on powers of 2 - if you want to try primes, you can (but have you *really* got a 50-share project to burn?) - and are:

$$3.14159 \sim \frac{2^{267377}}{10^{80488}} \qquad 1.414213 \sim \frac{2^{3216082}}{10^{468137}}$$

(I haven't the faintest idea what the 3216082nd prime number is.†) The final example picks up the Biblical theme to be found elsewhere in this journal and is given by the opening digits of 2^{1285}, which needs some shrewdness to evaluate.

References

[1] Arnol'd, V. I., *Mathematical Methods of Classical Mechanics*, pp. 71, 72 §D, E.
[2] Baker, A., *A Concise Introduction to the Theory of Numbers*, p. 43.
[3] Baker, A., op. cit. p. 4.
[4] Hlawka, E., *The Theory of Uniform Distribution.*

† Dr. Pinch has: he kindly informed me it is 53820209 - Ed.

Mathematical Limericks

Thomas Bending and Mark Wainwright
report on recent work at the
Cambridge University Mathematical Limerick Laboratory (CUMLL)

This year has been an exciting one for the Laboratory, seeing the completion of our initial research phase and the production of our first prototypical limericks. After much patient investigation we have discovered that there seem to be two approaches to the basic problem of constructing Mathematical Limericks (MLs), both yielding partial success.

Basic Principles

A. Metre We have at all times worked with the strictest metrical conditions, for it is the metre which makes or breaks a limerick.

A limerick comprises five lines with 3, 3, 2, 2, 3 feet respectively. Of first importance is the anapestic (◡◡–), dactylic (–◡◡) or amphibrachic (◡–◡) construction (where – represents a stressed and ◡ an unstressed syllable). The conditions for a limerick are stronger than for classical poetry because of the requirement that the metre (viz. a stressed syllable alternating with two unstressed syllables) must continue throughout the limerick, even across the line breaks. Thus, while it is possible (and indeed usual) to change between the three basic rhythms, this may be done only with great technical care, giving rise to complex rules in which, for example, a choriamb (–◡◡–) must be substituted in changing from a dactylic to an anapestic scheme.

B. Rhyme Lines $\{1,2,5\}$ rhyme and lines $\{3,4\}$ rhyme. It is desirable that lines 1 and 3 do not rhyme. Rhyme is taken in the strict sense. Thus two lines l_1, l_2 rhyme if
(i) the syllables after the last stressed syllable (S_i in line l_i) are identical, and
(ii) S_1, S_2 begin with distinct consonant phonemes but are otherwise identical.

Thus "reduction" and "construction" are strict rhymes but "compulsion" and "repulsion" are not. "Ease" and "p's" are not strict rhymes, but are acceptable if "ease" is preceded by a word ending in a consonantal phoneme distinct from a hard p, e.g.

...He multiplied p's / With the greatest of ease...

C. Meaning While A and B above apply to all limericks, the predominant characteristic of a mathematical limerick is the method by which it is expressed. It should be writable in symbols which are largely mathematical, or at least have mathematical meanings. Ideally it should have inherent meaning and, better still, be true. The following demonstrates this condition neatly:

$$[(k^2+3=3d)\wedge(a=13+\lfloor e\rfloor)\wedge(d=3a/50)]\Rightarrow k=\sqrt{-0.3}. \tag{1}$$

Astute readers will perceive that it is to be read

If k squared plus three is three d (2)
And if a is thirteen plus int e
Then if d is three a

Over fifty then k

Is the square root of minus point three.

D. Uniqueness The above limerick also fulfils the fourth condition for an optimal ML, i.e. that a competent reader should be able, by looking at the mathematical formulation, to deduce the limerick form, which ideally is unique. That is not to say that this need be possible without application and ingenuity on the part of the reader.

Approaching the Problem

Apparently there are two approaches to the problem of an ML. It seems that it is possible to concentrate either on conditions A and B above (the technical approach) or on condition C (the semantic approach). Both have been tried at CUMLL.

TB, using the latter approach, worked on the following:

If ϕ is a continuous function on a closed interval in $\mathbf{R}$ bounded by K', and $\psi = \phi^{-1}$, then $\psi(K')$ exists. (3)

If T, U, V are topological spaces, and $T \xrightarrow{p} U \xrightarrow{q} V$ with p, q continuous surjections, and V is compact, then T, U are compact. (4)

Clearly neither of these defines a unique, well-formed limerick. However a straightforward application of the Limerick Existence Theorem (LET) shows that each has at least one solution. After much research, TB discovered the following which, while satisfying A and B above, still sound rather strained:

The map phi has K prime as a bound, (5)
A domain that's without brackets round,
Is continuous and psi
Is the inverse of phi
Therefore psi of K prime can be found.

If the map q maps U onto V (6)
And if V is compact and if p
Is continuous like q
And maps T onto U
Then U must be compact, as must T.

The technical approach has been investigated extensively by MW. This has so far had the interesting but undesirable tendency to produce limericks which, while they are technically excellent and even satisfy condition D to some extent, fail completely on condition C: that is to say, they have been entirely devoid of actual meaning. Thus the following limerick lemmata have been produced:

$$\text{If } y = \eta \cos\zeta \text{ and if } x = \theta \sin\phi, \text{ then } \left.\frac{\partial y}{\partial \zeta}\right|_{r,\theta} = \frac{1}{\pi} \qquad (7)$$

$$\int_0^{\frac{\pi}{3}} \beta\, \mathrm{dc} = \frac{\partial^2}{\partial\eta\partial\nu}(|g|) \tag{8}$$

The reader is urged to spend some time in searching for the limerick forms of these statements. Canonical forms are given at the end of this article.

Some Further Pure Research Results

1. There is exactly one non-trivial English rhyme in $\mathbb{Z}$, viz. "7" and "11".

Note that a number of other languages contain at least one rhyme, e.g. in German 2 (zwei) rhymes with 3 (drei), while in French 6 (six) rhymes with 10 (dix), and so on - much work remains to be done in this area. There is another English rhyme in $\mathbb{R}$: e, when preceded by a suitable consonant, rhymes with 3.

2. The use of algebra vastly increases the number of rhyming mathematical entities available.

Rhyme in fact partitions the English alphabet $\mathbb{A}$ into 15 equivalence classes, four of them non-trivial. If we include $\mathbb{G}$ (the Greek alphabet), we obtain the following non-trivial equivalence classes:

$\mathbb{A}$	a j k	b c d e g p t v	i y	q u	n			
$\mathbb{G}$			$\xi\ \pi\ \phi\ \chi\ \psi$	$\mu\ \nu$		$\beta\ \eta\ \zeta\ \theta$	τ	
$\mathbb{Z}$		3		2	10		4	7 11

The Future

MW has made some progress towards unifying the above two approaches, which is hoped will turn out to be logically equivalent - this is the Extended Wainwright-Bending Hypothesis. Applying the technical approach to (3) yields partial success, although it should be noted that S_2 and S_5 are not distinct:

If phi is continuous on J (9)
Which is closed and is bounded by K
And the inverse of phi
Is denoted by psi
We have psi is defined upon K

A question which requires further work is the status of such words as "continuous". In (3) and (4) above it is treated as having three syllables, whereas in fact it has four. At present such matters appear to be best left to individual authors.

Finally we mention the Four Colour Project, the aim of which is to express the Four Colour Theorem in a limerick. Calculations based on the proof of the LET indicate that finding a solution would require many hours of computer time, in particular the checking of up to 1,936 possible rhyme schemes. Such a solution may not prove acceptable to many mathematicians, but we would be very interested to hear of any progress in this direction.

Appendix

If y equals eta cos xi
And if x equals theta sin phi
Then d y by d zeta
At constant r theta
Is equal to one over pi.

The integral beta d c
Between zero and pi upon three
Equals partial d two
By d eta d nu
Of the absolute value of g.

Archimedean Poetry

In the pages of Eureka, the Archimedeans have made many major (if usually anonymous) contributions to English literature. The following is but a brief selection, from issues 14, 15 and 7:

Lebesgue Measure
Is a vague pleasure
But the sets of Borel
Are absolute hell.

Electricity
Makes me write verse of greater or less felicity
And I am filled with the inspiration of the Muses
When I think of Laplace's equation or mend the fuses.

Little beast one doesn't smile on, almost smaller than ϵ,
(I am, of course, referring to that pest, the common flea)
It hops around quite happily in very large parabolae
And its maximum displacement equals v^2 *over* g.

In issue 1, Mr. Hope-Jones of Eton College contributed the first verse of a proposed Society anthem, sung to the tune of Hymn 341, Ancient and Modern:

All praise to Archimedes, who weighed the royal hat.
Displacing quarts of h. and c., upon the bathroom mat.
For that unending decimal, we mortals know as π,
He found three-and-one-seventh, was just a bit too high.

Not the Schedules for the Mathematical Tripos, 1988-9

Candidates should be familiar with none of the following:

Algebra I
"A one by one matrix has one column and one row, and the same number in both"
"I don't really understand the summation convention"
Vector Calculus
"If you've got a problem with this [the summation convention] then go back, write the whole thing out using sigma notation and convince yourself that it's better not to have problems"
A mnemonic for the vector triple product $\underline{a} \times (\underline{b} \times \underline{c}) = (\underline{a} \cdot \underline{c})\underline{b} - (\underline{a} \cdot \underline{b})\underline{c}$:
"Take inner product of outer two vectors and multiply by third and then subtract a term to make result antisymmetric under interchange of the two vectors bracketed together"
Algebra II
"One property we do know very well happens: $a + b = b + c$" *(For general* $a, b, c \in \mathbb{Z}$*)*
"g inverse is called an inverse to g"
Analysis II
"Clearly $x < y$ if and only if $y < x$"
"This series doesn't converge for the grossest of reasons"
Statistics
"Let the pendulum swing 100 times, count the number of swings and divide by 100"
"One quickly loses one's conscience doing statistics"
"We're not doing mathematics; this is statistics"
Mathematical Methods
"Of course, ω_k may not be real - it may be negative"
Quantum Mechanics
"This is a relativistic situation, because light moves at the speed of light"
"Just because they are called "forbidden" transitions does not mean that they are forbidden. They are less allowed than "allowed" ones, if you see what I mean." *(No - Ed.)*
Rings
"I must confess that every previous time I've proved this on a blackboard in public, I've got it wrong"
Fluid Dynamics
"You musn't be too rigid when doing Fluid mechanics"
Geometry
"I'll just define it in a more complicated way"
"I'm fudging one or two things here, but it's a short course"
Relativistic Electrodynamics
"There are some bits at the end of the course I don't really understand, but the students don't normally get that far"
Measure Theory
"f is as unique as you could reasonably expect it to be"
"I'll just remind you, if you haven't seen this before"
"More or less without too much loss of generality" (MOLWTMLOG)

"You've got to check it - and when I say you, I mean you, not me"

Algebraic Topology

"The n-simplex has all properties beginning with 'c'"

Riemann Surfaces

After getting in a mess: "It's not my fault – it's mathematics"

Differential Analysis and Geometry

"This is technically a confusion"

Set Theory

"I shall be handing out a... a... an aide-memoire"

"If it looks like a duck, quacks like a duck and swims like a duck, then it is a duck. Hence by transfinite induction, $\sigma(\alpha, \beta) = \alpha + \beta$"

"The axioms of set theory are like binary nerve gas - you put them together and run like hell"

Galois Theory

"This is very much like 'A'-level Physics"

Algebraic Geometry

"So there are 16 singular points - 12 at the bottom and 3 at the top"

Groups

"This is really just group theory in a particularly well-understood topos, but don't write that down or I'll get the sack"

"Not so much a double coset table, more a pile of junk"

Number Theory

"So this is a prime. I don't think it's the largest one"

"Of course this is true for more general values of 5"

"117 is prime, so this is easy"

"This is an integer, so it's a positive integer"

Linear Analysis

After a question defeating a certain member of the audience, "What? A Greek fisherman who does not know what an ϵ-net is?"

Convex Optimisation

"It's supposed to be called the "posynomial". Where they got that from I do not know – it took courage to write that down."

There are no formal schedules for Part III, but the attention of candidates is drawn away from the following:

"Damn! I'm running out of integers!"

"All numbers are totally irrelevant, unless you're doing Astrophysics"

From a substitute lecturer: "Good morning. For those of you who don't know me, I am not Dr. X. I am Dr. X's representative on Earth."

"Perhaps it would be best if this argument remained a deep mystery to you."

Legal disclaimer: Not all of the above theory was lectured at Cambridge, and not all of it by this year's lecturers! My thanks to Ian Redfern, Bob Dowling, Ruth Lilley, Tim Auckland and others for their diligent taking of lecture notes.

Neutron

Helena Verrill

Neutron (as described here) is a board game, invented by Robert A. Kraus. It first appeared in Games and Puzzles magazine [1], and was introduced to the Puzzles and Games Ring (in a slightly modified form) by Mark Owen a couple of years ago.

How to Play

The following instructions are written so as to ensure that even the most obstinate reader is unable to misinterpret them. I may not have succeeded; I expect that those who try hard enough will find my instructions ambiguous, but if so, just make something up.

Neutron is played by two players on a five-by-five board with eleven pieces: five for each player, and a "Neutron". The players are called White and Black, and their pieces similarly, as usual.

These start as shown in Figure 1.1, with each player's pieces initially on their back ranks, and the Neutron in the middle.

The aim is to win, which will happen when your opponent loses. You lose when you are unable to move, or if the Neutron is on your back rank; the game is symmetric.

A draw can be achieved by infinite play with neither player losing. This could be agreed upon, since there are not an infinite number of positions. Actually, a draw seems a very likely outcome of perfect play.

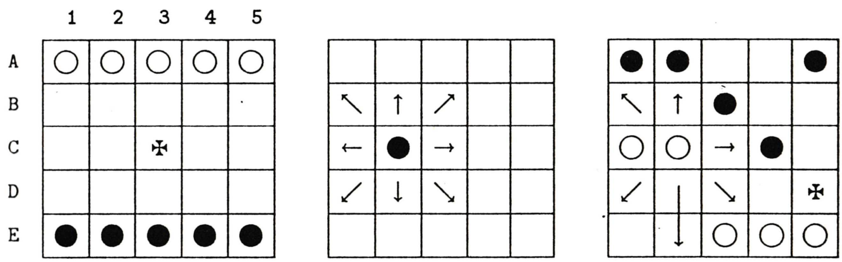
Players take alternate turns. A player's turn consists of a move of the Neutron, then a move of any one of that player's pieces, in the allowed way. The pieces all move in the same fashion; decide on the piece to be moved and the direction in which to move it. The allowed directions are shown in Figure 1.2. The piece must be moved in a straight line as far as possible in the chosen direction, subject to the confines of the board, and the condition that no two pieces may be on the same square at the same time, or pass over each other. The piece must actually move, i.e. it can't just remain where it is by moving as far as possible in an impossible direction. An example showing the possible moves of a particular piece is shown in Figure 1.3.

Notation

In the following I shall denote a move by eight characters, the initial pair for the position of the Neutron before it is moved, the next pair for the position after, then the positions of the piece moved before and after. Also, I shall take White to move first (i.e. Black to start), but I don't think there is any general convention.

Some Examples and Notes

Although Neutron is not particularly complicated, and games can be very short, it is not entirely trivial, and requires looking ahead a reasonable distance, which is difficult since there is a lot of branching. Computers should play Neutron better than humans, really, and Colin Wright has written a program which probably does. However, a lot of computer Neutrons I have played don't seem to be very good.



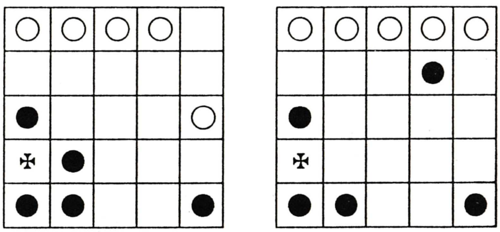
Figures 1.1, 1.2 and 1.3

The position in Figure 2.1 is very strong for Black (the positions around the Neutron being important, the others arbitrary). However, there are exceptions, and this kind of configuration does not always ensure a win for Black.

A good tactic in general seems to be to move the Neutron to any of `B1`, `B5`, `D1` or `D5` on your go. Dispensing with Archimedean tradition, I shall call these squares ★ squares, and the above sequence a ★ trap. It's also a good idea to restrict your opponent's options, since at least it's easier to look forward then.

If you don't want to win so much, defensive play (without attack) can lead to quite long games. For example, consider the sequence `C3B3E3C3`, `B3D5A3C5`, `D5D1C3D2` followed by the four moves `D1B1C5C1`, `B1B5D2B4`, `B5D5C1C5`, `D5D1B4D2` repeated for as long as you want to play. The player deviating from this sequence seems to be at an initial disadvantage, but I don't know whether or not deviation is losing.

Having too few pieces on your back rank is risky, but having all your pieces on your back rank can be equally dangerous, e.g. see figure 2.2, which is a loss for White.



Figures 2.1 and 2.2

An Example Analysed

The following example is interesting because the initial response by White looks quite reasonable; it looks as if it is going in the right direction for a ★ trap. However, after a little thought it becomes apparent that it loses. The following is only a sketch of the analysis, but filling in the gaps is easy; there is no space for a full painstaking analysis.

Numbering

The initial two moves are given; the position after these moves is shown in Figure 4.1. Branching occurs on most of White's following moves, but not on Black's, since I only need to find one winning response to show that Black wins. Figure 3 represents the branches of the game. The nodes represent positions and the lines between the nodes represent moves.

The move numbers give the number of the move as played, then a dot to indicate branching, and the subsequent numbers giving the branch number.

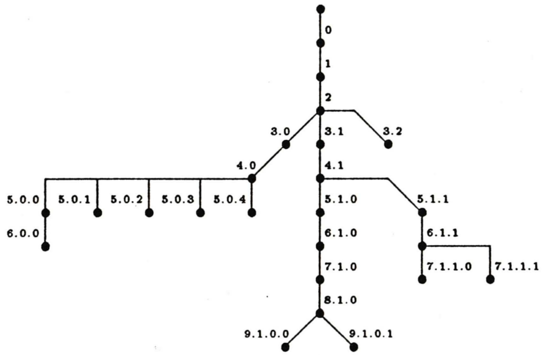


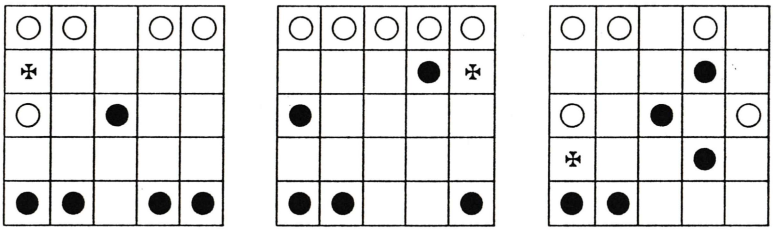
Figure 3

0	C3B3E3C3	
1	B3B1A3C1	This is the losing move; once Black plays:
2	B1B5E4B4	nothing can be done to avoid a loss.
		Note the threat of Black moving the Neutron to either
		D1, D2, B3 on his next move.
3.0	B5D5C1A3	Consider this response to begin with,
		which looks fairly safe, covering the back rank.
4.0	D5B5C3C1	This threatens B5D1B4E2. There are several
		possibilities to consider now. This is Figure 4.2.
5.0.0	B5D5A5C5	Stops return of Neutron to B5, but allows
6.0.0	D5D1B4D2	★ trap, which is pretty good for Black.
		This position is shown in Figure 2.1.
		We may, for example, get sequences like
		D1B3C5C2 to block B3D1, but loses anyway to:
		B3B1E5E3 Note that E5E3 could be replaced by

		virtually anything else here. The win for Black is obvious.
	B1B5....	This is forced, and White is unable to block both A4 and A5.

So try something else. Still considering Neutron to D5. Can't block with A2D2 (because of D5A2) so must be A1D4 or A4D1 (if I was playing, I think I'd prefer the latter - but it's lost anyway, so I probably wouldn't be playing).

5.0.1	B5D5A4D1	What about this, then?
6.0.1	D5B5E5C5	
7.0.1	B5D3A5B5	So as not to lose immediately to D3B5E2C4.
8.0.1	D3B1C5C2	This wins anyway, followed by something like
9.0.1	B1B3B5A4	
10.0.1	B3B1B4B2	Now Black has won.
5.0.2	B5D5A1D4	Try the other one then.
6.0.2	D5B5E5C5	
7.0.2	B5D3D4C4	Blocks D3B5, and the possibility of D3B1C5C2.
8.0.2	B3D1B4D2	The ★ situation again, so very threatening; also B5.
9.0.2	D1B3C4C2	
10.0.2	B3B5E2C4	Now White has obviously lost, not being able to cover back rank.
11.0.2	B5B1....	No way of blocking both A1 and A2.



Figures 4.1, 4.2 and 4.3

Consider Neutron to D3 at move five. Then must have either A2D2 or A4D1.

5.0.3	B5D3A2D2	
6.0.3	D3B5E5C5	
7.0.3	B5D3A5B5	This is 'forced', since otherwise Black plays D3B5E2C4.
8.0.3	D3B1C5C2	
9.0.3	B1B3....	Neutron move is forced, and can't block both A2 and A3.
5.0.4	B5D3A4D1	

6.0.4	D3B5E5C5	
7.0.4	B5D3A5B5	Must play this so as not to allow D3B5E2C4.
8.0.4	D3B1C5C2	
9.0.4	B1B3B5A4	Must play B5A4, so as not to lose to B3A4.
10.0.4	B3B1B4B2	Black wins.
		Other plays along these lines lose similarly.
		Consider something else at move three, then:
3.1	B5D5A5C5	But this doesn't help the D5D1 threat.
4.1	D5D1E5D4	Position after this move shown in Figure 4.3.
5.1.0	D1D3	If this is going to be Neutron move, must block D3B5E2C4, say with
5.1.0	D1D3C5C4	(otherwise, could block with D1D3A4B5, say, but then D3D1D4D2, D1A4... White loses.
6.1.0	D3D1D4D2	
7.1.0	D1B3C1A3	Otherwise can't block B3D1C3C2 - a clear loss.
8.1.0	B3D1C3C1	But then this looks pretty lost for White anyway.
9.1.0.0	D1B3C4C2	
10.1.0.0	B3B1B4B2	Wins for Black.
		Or, for another way of losing from 7.1.0, try
9.1.0.1	D1B3C4D3	Just for the sake of considering not C4C2, really.
10.1.0.1	B3D1C4C3	
11.1.0.1	D1B3....	Need to stop B3D1C3C2. Now an obvious loss for White.
5.1.1	D1B3C1A3	
6.1.1	B3D1C3C1	
7.1.1.0	D1D3....	Need to block D3B5E2C4, but then can't stop D3D1D4D2
7.1.1.1	D1B3....	That leaves B3D1D4D2 unblocked, so White loses, but can only block with A2D2, allowing B3A2; or block with C5C2, allowing B3B1B4B2.
3.2	B5D3....	Considering this as White's Neutron move three leads also to a loss, by similar considerations. Also, there is an extra threat for Black to play D3B1 to his advantage.

So a reasonable-looking response has been found to be losing. Can you find other such situations?

Neutron is easy and fun to play and analyse. I hope you find it so, too.

Appendix: Classical Neutron

"Classical" Neutron (as distinct from "Cambridge" Neutron, described above) is the original form described in [1] and later amended in [2] by the inventor. In [2], White begins, but moves only a piece, not the Neutron; play continues normally except that the victory is achieved by moving the Neutron to a player's *own* back rank.

References

[1] "Playroom", *Games and Puzzles 71* (1978).
[2] "Playroom", *Games and Puzzles 73* (1979).

Back to Square One

Derek J. Colwell, John R. Gillett and Brian C. Jones
North Staffordshire Polytechnic

Many board games involve traversals of a circuit of squares. Suppose movement around such a circuit is controlled by an M-sided die, which is not necessarily unbiased. Then this paper investigates the expected number of throws of the die which are needed to return to the starting square.

In board games such as *Monopoly* a circuit of squares has to be repeatedly traversed, with a competitor's movement around the circuit being controlled by the score obtained by throwing an unbiased die. The question we shall investigate in this paper is concerned with a competitor's expectation of returning to a given square.

To introduce more precision into this question suppose the circuit consists of N squares, numbered $1, 2, ..., N$. Then, starting from square r $(1 \leq r \leq N)$, what is the number of throws of the die a competitor can expect to make before landing on square r again?

Initially we considered this question for the usual unbiased 6-sided die. However, somewhat to our surprise, we have discovered that our result does not depend on the die being unbiased or 6-sided. We shall therefore assume that the die is M-sided and that the probability of obtaining each of the values $1, 2, ..., M$ is non-zero and independent of the square on which the competitor is positioned. This led us to proving the following theorem.

Theorem *A board game consists of a circuit of N squares numbered $1, 2, ..., N$ and a competitor's movement around this circuit is controlled by an M-sided die, which is not necessarily unbiased and which can take each of the values $1, 2, ..., M$ with a non-zero probability. The expected number of throws of the die which are needed to return again to the starting square is N.*

Proof Without loss of generality suppose we start from square 1. Then the theorem is trivially true for a 1-sided die. Suppose that $M \geq 2$ and that $P = (p_{ij})_{N \times N}$ is the probability transition matrix of the movements of a competitor around the squares, so that p_{ij} is the probability of moving from square i to square j with one throw of the die. Then

$$\sum_{j=1}^{N} p_{ij} = 1 \quad (i = 1, 2, ..., N). \tag{1}$$

Let $\underline{p}_i$ be the i-th row vector of P. Then in terms of the elements of $\underline{p}_1$,

$$\underline{p}_i = (p_{1,N-i+2}, p_{1,N-i+3}, ..., p_{1,N}, p_{1,1}, ..., p_{1,N-i+1}), \qquad (i = 2, ..., N).$$

Thus the elements of the vector $(p_{1j}, p_{2j}, ..., p_{Nj})$, for each j $(1 \leq j \leq N)$, are a permutation of the elements of p_1. Hence

$$\sum_{i=1}^{N} p_{ij} = 1 \quad (j = 1, ..., N). \tag{2}$$

For example, if $M = 2$, $N = 4$ and the die takes the values 1 and 2 with probabilities $\frac{1}{4}$ and $\frac{3}{4}$ respectively,

$$P = \begin{pmatrix} 0 & 1/4 & 3/4 & 0 \\ 0 & 0 & 1/4 & 3/4 \\ 3/4 & 0 & 0 & 1/4 \\ 1/4 & 3/4 & 0 & 0 \end{pmatrix}.$$

Suppose that a competitor has reached square i $(1 \leq i \leq N)$ and that the expected number of throws needed to reach square 1 from this position is denoted by m_i. Then

$$\begin{aligned} m_i &= 1 \times p_{i1} + \sum_{k=2}^{N} p_{ik}(1 + m_k) \\ &= 1 + \sum_{k=2}^{N} p_{ik} m_k. \end{aligned} \tag{3}$$

Writing

$$\underline{m} = (m_2, m_3, ..., m_N)^T, \quad \underline{\ell}_{(N-1)\times 1} = (1, 1, ..., 1)^T$$

and

$$Q = (q_{ij})_{(N-1)\times(N-1)}$$

where $q_{ij} = p_{i+1,j+1}$, then

$$\mathrm{I}\underline{m} = \underline{\ell} + Q\underline{m}$$

or

$$(\mathrm{I} - Q)\underline{m} = \underline{\ell}. \tag{4}$$

However, before $\underline{m}$ can be obtained from this equation it is first necessary to show that $(\mathrm{I} - Q)^{-1}$ exists or, equivalently, that $Q^k \to 0$ as $k \to \infty$, [1].

This will be demonstrated by introducing a matrix

$$T = \begin{pmatrix} 1 & \underline{0}_{1\times(N-1)} \\ p_{21} & \\ \vdots & Q \\ p_{N1} & \end{pmatrix}.$$

This matrix is the probability transition matrix of a competitor moving around the squares in such a way that having landed on square 1, he stays there. That is, square 1 is an absorbing state.

It follows that

$$T^k = \left(t_{ij}^{(k)}\right)_{N\times N} = \begin{pmatrix} 1 & \underline{0}_{1\times(N-1)} \\ t_{21}^{(k)} & \\ \vdots & Q^k \\ t_{N1}^{(k)} & \end{pmatrix}$$

is the probability transition matrix of the movements of a competitor after k throws of the die.

As a competitor circuits the board, the probability of landing on square 1 on each circuit is at least as great as the probability of the least likely value on tossing the M-sided die. Since this value is non-zero it follows that the probability of landing on square 1 becomes certain, so that $t_{i1}^{(k)} \to 1$ as $k \to \infty$ $(i = 1, 2, ..., N)$. But, since T^k $(k = 1, 2, ...)$ is also a probability transition matrix,

$$\sum_{j=1}^{N} t_{ij}^{(k)} = 1 \quad (i = 1, 2, ..., N).$$

Hence

$$t_{ij}^{(k)} \to 0 \text{ as } k \to \infty \quad (2 \leq i, j \leq N),$$

and thus, as required,

$$Q^{(k)} \to 0 \text{ as } k \to \infty. \tag{5}$$

Setting

$$\mathrm{I} - Q = A = (a_{ij})_{(N-1)\times(N-1)}$$

and using (5), it may be deduced that

$$A^{-1} = B = (b_{ij})_{(N-1)\times(N-1)}$$

exists. Hence it follows from (4) that

$$m_{k+1} = \sum_{j=1}^{N-1} b_{kj} \quad (k = 1, 2, ..., N-1) \tag{6}$$

and, from (2), that

$$\sum_{i=1}^{N-1} a_{ik} = 1 - \sum_{i=1}^{N-1} p_{i+1,k+1} = p_{1,k+1} \quad (k = 1, 2, ..., N-1). \tag{7}$$

The theorem now follows from (3), (6) and (7), since

$$
\begin{aligned}
m_1 &= 1 + \sum_{k=2}^{N} p_{ik} m_k \\
&= 1 + \sum_{k=1}^{N-1} p_{1,k+1} m_{k+1} \\
&= 1 + \sum_{k=1}^{N-1} \sum_{i=1}^{N-1} a_{ik} \sum_{j=1}^{N-1} b_{kj} \\
&= 1 + \sum_{k=1}^{N-1} \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} a_{ik} b_{kj} \\
&= 1 + \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \sum_{k=1}^{N-1} a_{ik} b_{kj} \\
&= 1 + \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \delta_{ij}, \quad \text{since } AB = \mathrm{I} \\
&= 1 + N - 1 \\
&= N.
\end{aligned}
$$

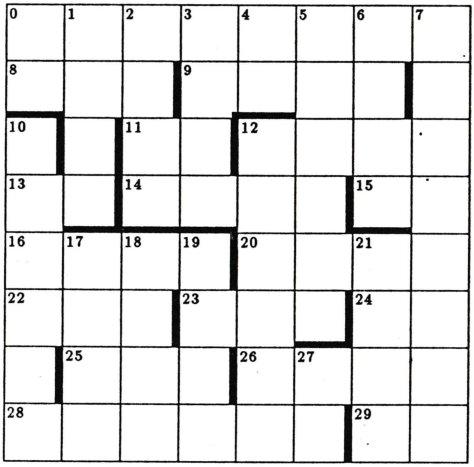
□

Reference

[1] L. Fox, *An Introduction to Numerical Linear Algebra*, Oxford, 1964.

Interstice

by Quinapalus



$p < q < r$ are 3-digit primes, fixed throughout. All answers are to be written in base 10, unless otherwise stated.

Across

0	pqr
8	Sum of two squares, in base 8
9	
11	2×13A
12	Multiple of p
13	
14	Multiple of r
15	Prime, written in base 8
16	$p + q + r$
20	
22	
23	3D÷8
24	
25	$(24\text{A})^2 + 2$
26	Square
28	Multiple of 81
29	$3 \times 0\text{D} - 2$

Down

0	
1	2×19D
2	19D ÷ 2
3	
4	$\sqrt[3]{20\text{A}} + \sqrt{9\text{A}}$
5	
6	2×17D
7	$2pqr$
10	(Power of 2)+p
12	
17	Square
18	1D+2D
19	
21	
27	Composite

Single Vertex Graphs

Clive Monk

The purpose of this article is to explain a method of associating functions with various objects in such a way that cutting up the objects corresponds to differentiating the functions, and sticking objects to other objects corresponds to multiplying their functions together. By an "object" we in fact mean an ordered pair (T, p) where T is a topological space which without loss of generality can be taken to be connected, and p is an element of T, but I've tried to write this article in such a way that most of it doesn't require a knowledge of what these terms mean. We abbreviate (T, p) to T_p, and where possible (which is just about everywhere) we draw a picture of the thing, representing p by a black dot. For the reader who knows about such things, virtually all the topological spaces that we consider are triangulable. With each object we associate precisely one function†. We represent this function by a picture of the object, and equations such as (24.1) are to be interpreted as relations between the functions rather than assertions that the objects at either side of the equals sign are "the same" (i.e. homeomorphic). Had we been eager to assert this, we would have written $\cong$ rather than $=$, but they aren't, so we weren't, so we didn't. We denote by L^n the single vertex graph with n edges, where n is a non-negative integer. Each edge is a loop, hence the notation. We call the vertex v and write L^1 as L. Different edges are not considered the same. This isn't as idiotic a remark as it sounds since often in questions about selecting different coloured balls from urns, for example, different balls are considered "the same" when they are the same colour. We shall be concerned with the number of ways of 'drawing' L^n on our object in such a way that v gets mapped to p, and the edges don't intersect themselves or one another except at p. This concept will be made more precise later. The idea is that two drawings are equivalent if and only if one can be "slid around" the object to get the other.

The first object we consider is the plane ($\mathbf{R}^2$). This is homeomorphic to an open disc, suggesting that $\odot$ is a sensible notation to use. How many ways are there of drawing L^n on the plane? We are not allowed to turn the plane over, and probably wouldn't be able to if we were anyway, because it's fairly big! Well, there is just one way when $n = 0$, and a picture of it can be found at the end of this sentence. For $n = 1$, there is one way also, while for $n = 2$ there are three ways: two which look like

 and one like 

What about larger n? We use generating functions to tackle this question. Denote by $\odot^{L^n}$ the answer for n, and put

$$T_p(x) = \sum_{n=0}^{\infty} T_p^{L^n} \cdot \frac{x^n}{n!} \tag{1}$$

where

$$n! = \prod_{r=1}^{n} r \qquad (n \geq 0) \tag{2}$$

† Strictly speaking, this isn't true, but this remark is more confusing than helpful at the moment, which is why it's a footnote.

and $T_p = \odot$. This defines the **generating function** $\odot$ (which for the moment is to be thought of as a formal power series) in terms of the $\odot^{L^n}$. We define the generating function T_p of any object T_p for which all $T_p^{L^n}$ are finite, by this equation. Then

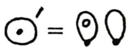
$$T_p^{L^n} = \left[\frac{\mathrm{d}^n}{\mathrm{d}x^n} T_p\right]_{x=0} \qquad (n \geq 0). \tag{3}$$

Each T_p mentioned in this article has a sufficiently simple generating function to enable us to obtain from (1) an expression for $T_p^{L^n}$ which could, if we so desired, be written using only the symbols $n, \Sigma, \Pi, +, -, \times, \div, (,), =, 0$ and 1, together with the symbols which appear at the bottom of the summation and product signs – hence the title (if you think about it). The expressions which we give are not of this form, but may be put in this form by using (2), writing powers as repeated products, and using suitable bracketing. Note that if, for a given T_p, we only require an expression for $T_p^{L^n}$ and are not bothered what functions of n appear in it, then our task is extremely simple – $T_p^{L^n}$ is itself such an expression! Stirling's formula states that

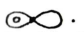
$$n! \sim e^{-n} n^n \sqrt{2\pi n}.$$

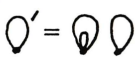
We use it each time we write down an asymptotic formula.

Now for the crucial idea. Replace your pencil with a pen-knife and draw the first edge. This splits the plane up into a  and a  , the hole in the middle of the former being due to the missing point at infinity. If we wanted to formalize this we would have to be more precise. In particular we would need to decide whether the above two objects should be taken to include their boundaries. This will be discussed later. Each of the remaining edges must be drawn on one of these objects in such a way that the image of v is unambiguous, ie is p. Likewise if, after differentiating a function, one gets a product fg of functions, then each further differentiation must be applied either to f or to g. The process is exactly analogous. In a nutshell,



(4)

Note that with  it does make a difference to the generating function if we change p. This is why our objects are ordered pairs. Note also that  multiplied by  is equal to . We adopt the convention that with objects such as the one on the left of (51.1), the black dots are to be identified to a single point p. Under this convention, equations such as (4) are unambiguous. Similarly



(5)

so that

$\int$  $dx = c + \ln$ 

(6)

Note that the right hand side of (6) is meaningful, since

 $(0) = 1$

(7)

Indeed, $T_p(0) = 1$ for any $T_p(x)$. This fact is surprisingly significant. In this article, we use it as a boundary condition whenever we solve a differential equation. We could draw

L^n on  *before* making the hole. There would then be $n+1$ ways of adding the hole. The same is true of . We write these facts down, (9) being for later use:

 $= \left(1 + x\frac{\mathrm{d}}{\mathrm{d}x}\right)$  (8)

 $= \left(1 + x\frac{\mathrm{d}}{\mathrm{d}x}\right)$  (9)

Similarly let  denote an arbitrary object which is as shown near p. Then

 $= \left(1 + 2x\frac{\mathrm{d}}{\mathrm{d}x}\right)$  (10)

It is not true in general that

 $= \left(1 + x\frac{\mathrm{d}}{\mathrm{d}x}\right)$ .

Consider  for example. It is not the case that  $^L = 2$  ^{L}. Infinity does not equal four. From (5) and (10) we have

 $= ($  $+ 2x$  $)$  (11)

(7) and (11) specify  completely since for $n > 0$ (11) gives  $^{L^n}$ in terms of its values for smaller n. So if we can find a solution of (7) and (11), then it is the correct solution. We write the claimed solution in the form of a complex valued function of a complex variable which is analytic in a neighbourhood of zero and then proceed to check it.

 $= \dfrac{1 - \sqrt{1-4x}}{2x}$ (12.1)

In this, and all other formulae which follow, it is intended that the value given by the Taylor expansion (when this converges) is the one taken. Despite the $2x$ on the bottom, zero is in fact a removable singularity and

 $(0) = \lim\limits_{x \to 0}$  $(x) = 1$

as required. The Taylor expansion is

 $= \displaystyle\sum_{n=0}^{\infty} \frac{(2n)!}{(n+1)!} \frac{x^n}{n!}$

and is valid for $|x| < \frac{1}{4}$. So it suffices to verify (11), which we now do, extracting equations (12.2) and (13.1) in the process. From (1) and (12.1)

 $^{L^n} = \dfrac{(2n)!}{(n+1)!} \quad (n \geq 0) \quad \sim \dfrac{\sqrt{2}}{n}\left(\dfrac{4n}{e}\right)^n$ (12.2)

Of course the reader may well have guessed this already by looking at small values of n. This is in fact how the author found (12.1) in the first place. From (10) and (12.1) we see that

 $$= \frac{1-\sqrt{1-4x}}{2x\sqrt{1-4x}} \qquad (13.1)$$

and hence that (5) is satisfied, thus verifying (11). Using (1) and (13.1),

 $$^{L^n} = \frac{(2n+1)!}{(n+1)!} \quad (n \geq 0) \quad \sim 2\sqrt{2}\left(\frac{4n}{e}\right)^n \qquad (13.2)$$

From (8) and (12.1), it follows that

 $$= \frac{1}{\sqrt{1-4x}} \qquad (14.1)$$

and again using (1) that

 $$^{L^N} = \frac{(2n)!}{n!} \quad (n \geq 0) \quad \sim \sqrt{2}\left(\frac{4n}{e}\right)^n \qquad (14.2)$$

At this point, it is clear that

 $=$   (15)

and combining this with (4) and (6) gives

 $= 1 + \ln$  (16.1)

Using (1) yet again (we use it without mention from now on) yields

 $$^{L^n} = \frac{(2n-1)!}{n!} \quad (n \geq 1) \quad \sim \frac{1}{\sqrt{2}n}\left(\frac{4n}{e}\right)^n \qquad (16.2)$$

What about the sphere S^2? We denote this by , the h (for "hollow") distinguishing it from the plane. It would be a mistake to say that  $=$  ^{2}, since although drawing the first edge does indeed give two objects which look like  , these are in fact indistinguishable. What we *can* say, however, is that

 $=$   (17)

Integrating gives

 $= \frac{1}{2} + \frac{1}{2}$  ^{2} (18)

Now we know, by subtracting the expression (10) gives for  from $2 \times (8)$, that

 $= 2$  $-$  (19.1)

$$^{L^n} = 2 \; ^{L^N} - \; ^{L^n} \qquad (n \geq 0) \tag{19.2}$$

Replacing one of the 's in (18) by the right hand side of (19.1) and using (4), (5) and (16.1) then yields

$$= \frac{1}{2}(3+x) - \frac{1}{2} \; + \ln \tag{20.1}$$

$$^{L^n} = \frac{(2n-1)!}{(n+1)!} \quad n \geq 2 \quad \sim \frac{1}{\sqrt{2}n^2}\left(\frac{4n}{e}\right)^n \tag{20.2}$$

It is possible to generalize (10). Let $= y_0$, $= y_1$, $= y_2$ and so on, and suppose that $y_0^{L^n}$ is finite for all n, so that we have a generating function for y_0 in the form of a formal power series. As before there are two ways of going from a picture of y_0 to a picture of y_r with L^n drawn on it. We could change y_0 to y_r and then draw on L^n, or we could start off with L^n drawn on and think about how to change y_0 to y_r. Thus

$$\sum_{r=0}^{\infty} y_r w^r = \sum_{n=0}^{\infty} \left(\sum_{i=0}^{\infty} w^i\right)^{2n+1} y_0^{L^n} \frac{x^n}{n!} \tag{21}$$

No algebra has been done here – (21) is just a mathematical way of writing this down. The nice way to think about (21) is that by summing the geometric series we can write it in the form

$$+ \; w + \; w^2 + \ldots = \frac{1}{1-w} \left(\frac{x}{(1-w)^2}\right) \tag{22}$$

We can then compare the coefficient of w^r on either side of this equation to obtain an expression for y_r in terms of y_0 for any r of our choice (exercise for reader). Taking $r = 1$ re-establishes (10) - this is the sense in which (22) is a generalization. It is well known that for m and r not both zero, the number of monotonically increasing functions from a totally ordered set with r elements to a totally ordered set with m elements is $\binom{m+r-1}{r}$. Indeed, there is a neat combinatorial proof of this which the reader may care to try and find. Using this idea with $m = 2n+1$ gives

$$y_r^{L^n} = \binom{2n+r}{r} y_0^{L^n} \qquad r, n \geq 0 \tag{23}$$

which is equivalent to (22).

From (12.1), (14.1) and (22) we obtain the following equations. (24) and (25) are needed later. (26) has been added just for the fun of it.

$$= \; = \frac{1}{(1-4x)^{\frac{3}{2}}} \tag{24.1}$$

$$^{L^n} = \; ^{L^n} = \frac{(2n+1)!}{n!} \quad n \geq 0 \sim 2\sqrt{2}n\left(\frac{4n}{e}\right)^n \tag{24.2}$$

$$= \frac{1+2x}{(1-4x)^{5/2}} \tag{25.1}$$

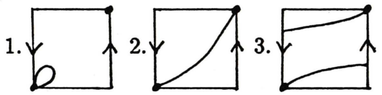

$$
L^n = \frac{(2n+2)!}{2n!} \quad n \geq 0 \sim 2\sqrt{2}n^2 \left(\frac{4n}{e}\right)^n \tag{25.2}
$$


$$
= \frac{1}{(1-4x)^{5/2}} \tag{26.1}
$$


$$
L^n = \frac{(2n+3)!}{6(n+1)!} \quad n \geq 0 \sim \frac{4\sqrt{2}}{3} n^2 \left(\frac{4n}{e}\right)^n \tag{26.2}
$$

Note that, for example, the second equality in (24.1) follows also from (13.1).

The Möbius band with p on its boundary is slightly harder. There are three ways of drawing on the first edge, and these are shown below:


(27)

Thus we get


(28)

which together with (10), (12.1), (13.1) and (24.1) gives us a first order L.D.E. which we can solve to get


$$
= \frac{1+\sqrt{1-4x}}{2(1-4x)} \tag{29.1}
$$

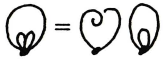

$$
L^n = \frac{1}{2}\left[4^n n! + \frac{(2n)!}{n!}\right] \quad n \geq 0 \sim \sqrt{\frac{\pi n}{2}} \left(\frac{4n}{e}\right)^n \tag{29.2}
$$

(10) then tells us that


$$
= \frac{1+4x+\sqrt{1-4x}}{2(1-4x)^2}. \tag{30.1}
$$


$$
L^n = \frac{2n+1}{2}\left[4^n n! + \frac{(2n)!}{n!}\right] \quad n \geq 0 \sim \sqrt{2\pi} n^{3/2} \left(\frac{4n}{e}\right)^n \tag{30.2}
$$

(13.1), (24.1) and (29.1) yield


(31)

a fact that will be useful later.

Despite our remark about not being able to write (8) in the same generality that we have written (10), there is nothing to stop us applying the operator $1 + x\frac{d}{dx}$ to various objects and seeing what we get. Indeed we *can* write equations such as

 $= \left(1 + x\frac{d}{dx}\right)$  (32)

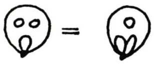
 $= \left(1 + x\frac{d}{dx}\right)$  (33)

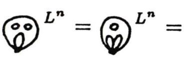
provided we allow p to wander around the boundary of the objects on the left (together with the slit, in (33)) any integral number of times. As with (8) the idea is that we can make the second hole before *or after* drawing on the graph. (14.1) and (32) give

 $= \dfrac{1-2x}{(1-4x)^{3/2}}$ (34.1)

 $L^n = (n+1)\dfrac{(2n)!}{n!} \quad n \geq 0 \sim \sqrt{2}n\left(\dfrac{4n}{e}\right)^n$ (34.2)

while (9), (24.1), (25.1) and (33) give

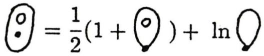
 $= \dfrac{1+2x}{(1-4x)^{5/2}}$ (35.1)

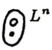
 $= \dfrac{(2n+2)!}{2n!} \quad n \geq 0 \sim 2\sqrt{2}n^2\left(\dfrac{4n}{e}\right)^n$ (35.2)

Now,

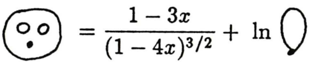
 (36)

and in view of (12.1), (14.1) and (34.1), we may integrate this to get

 (37.1)

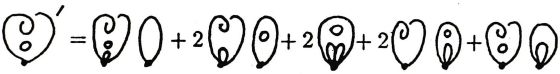
 $L^n = (n+1)\dfrac{(2n-1)!}{n!} \quad n \geq 1 \sim \dfrac{1}{\sqrt{2}}\left(\dfrac{4n}{e}\right)^n$ (37.2)

Arguing as in (32) and (33) tells us that

 (38.1)

 $L^n = (n+1)^2\dfrac{(2n-1)!}{n!} \quad n \geq 1 \sim \dfrac{n}{\sqrt{2}}\left(\dfrac{4n}{e}\right)^n$ (38.2)

where p is allowed to wander around the left hand side. (8), (9), (32) and (33) suggest that $1 + x\frac{\mathrm{d}}{\mathrm{d}x}$ punches holes in things, but we have to be careful here. Consider the Möbius band with a hole in it, where, as with the object on the left of (32), p is allowed to wander around its boundary (an integral number of times). Here our trick of adding the hole after L^n has been drawn won't work. For $n = 1$ there are indeed, as it happens, two ways of adding the hole to (27.2). In (27.1) and (27.3), however, there are three ways. It is left as an exercise to the reader to spot what these ways are. All is not lost, however, since we do at least have

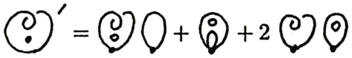
 (39)

and from this equation, together with (12.1), (13.1), (14.1), (24.1), (25.1), (29.1), (30.1) and the idea behind (10), we can form a first order L.D.E. as before and solve it to get

$$= \frac{4x + (1-2x)\sqrt{1-4x}}{(1-4x)^2} \tag{40.1}$$

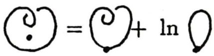
$$^{L^n} = n4^n n! + (n+1)\frac{(2n)!}{n!} \quad n \geq 0 \sim \sqrt{2\pi} n^{3/2} \left(\frac{4n}{e}\right)^n \tag{40.2}$$

Let us now consider a Möbius band with p somewhere in the middle, and not allowed to wander around. Since a Möbius band is homeomorphic to a projective plane with a hole cut out of it, we have



(41)

A single integration together with (12.1), (14.1), (24.1), (29.1) and (40.1) then gives us

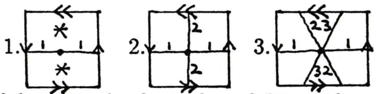


(42.1)

$$^{L^n} = {}^{L^n} + {}^{L^n} = \frac{4^n n!}{2} + (n+1)\frac{(2n-1)!}{n!} \quad n \geq 1 \sim \sqrt{\frac{\pi n}{2}} \left(\frac{4n}{e}\right)^n \tag{42.2}$$

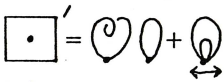
It is interesting to compare (42.1) and (37.1).

The reader may wonder what the generating function is for the projective plane itself. We consider this question for the case where p is fixed.



(43)

(43.1) shows a way of drawing the first edge of L^n on the projective plane. There are only two ways - this way, and the way in which it appears as a small loop. In the way shown, the two asterisks are indistinguishable, perhaps suggesting that an equation like (17) is called for. With the projective plane, however, one can go on adding edges for as long as one pleases in such a way that this symmetry is preserved (see (43.2), (43.3)). We get round this problem by observing that



(44)

where the double pointed arrow indicates that here turning over *is* allowed. Now

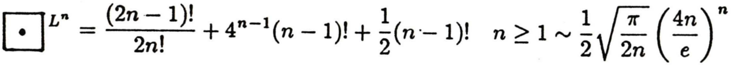
$$= \frac{1}{2}\left[+ \sum_{n=0}^{\infty} n! \frac{x^n}{n!}\right] = \frac{1-\sqrt{1-4x}}{4x\sqrt{1-4x}} + \frac{1}{2(1-x)} \tag{45}$$

and while we're at it

$$= \frac{1}{2}\left[+ \sum_{n=0}^{\infty} (n+1)n! \frac{x^n}{n!}\right] = \frac{1}{2(1-4x)^{3/2}} + \frac{1}{2(1-x)^2} \tag{46}$$

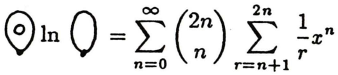
(12.1), (29.1) and (45) enable us to integrate (44). The result we obtain is

(47.1)

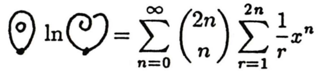


(47.2)

This seems like a good place to mention two rather pretty equations, the proof of which is left to the reader.



(48.1)



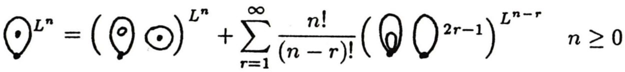
(48.2)

Comparing coefficients in (48.1) yields a result perhaps worth a mention,

$$\sum_{r=1}^{n} \frac{\binom{n}{r}^2}{2r\binom{2n}{2r}} = \sum_{r=n+1}^{2n} \frac{1}{r} \tag{49}$$

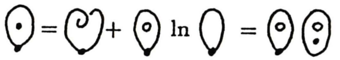
It is a well-known fact that the right hand side of (49) tends to ln 2 as $n \to \infty$. This is useful in obtaining the asymptotic formulae which follow.

The next object we tackle is  . There are two possibilities. Either there are no edges going from the outer point to the inner point, or there are $r \geq 1$ edges. This tells us that



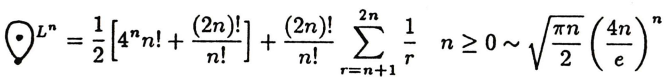
(50)

where $\frac{1}{r!}$ is taken as zero for $r < 0$. Multiplying by $\frac{x^n}{n!}$, summing over all n, and using (12.1), (13.1), (14.1), (16.1), (29.1) and (37.1) yields



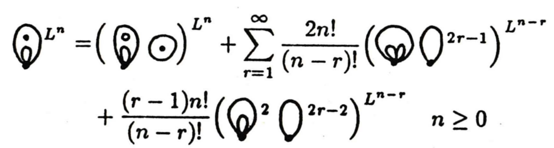
(51.1)

From (48.1), we then see that



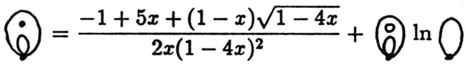
(51.2)

To give another example of this kind of reasoning,

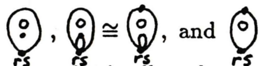


(52)

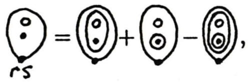
As before, we multiply by $\frac{x^n}{n!}$ and sum over all n. Use of (12.1), (13.1), (16.1) and (24.1) then gives

 (53)

We can calculate  $^{L^n}$ from (53) by using (48.1) (exercise for reader). What if we punch a hole in the left hand side of (51.1) and allow the outer point to wander around the boundary so as to get rid of any tangles? As with (40.1) we can't just apply the "hole punch" operator to the equation in which our object appears without a hole (in this case, (51.1)). We can, however, 'count' the number of edges going from the outer to the inner point, as we did in (50) and (52). For this method of attack, we need the functions associated with



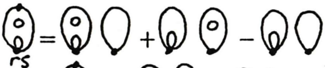
where in the first picture the outer point is allowed to wander around the boundary, and the 'rs', standing for 'return same', indicates that edges are not allowed to go from one point to the other in the pictures. Now



by which we mean

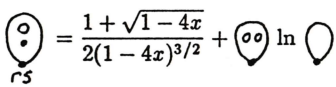
 (54)

Likewise

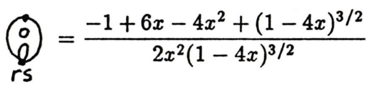
 (55)

 (56)

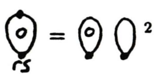
Using (14.1), (16.1), (34.1), (37.1) and (54) we get

 (57)

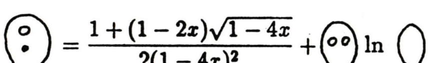
Using (12.1), (13.1), (14.1), (24.1) and (55) we get

 (58)

Using (12.1), (14.1) and (56) we get

 (59)

(12.1), (13.1), (14.1), (57), (58) and (59) provide us with the result:

 (60.1)

and hence from (48.1)


$$^{L^n} = \frac{1}{2}\left[n4^n n! + (n+2)\frac{(2n)!}{n!}\right] + (n+1)\frac{(2n)!}{n!}\sum_{r=n+1}^{2n}\frac{1}{r} \quad n \geq 0 \sim \sqrt{\frac{\pi}{2}}n^{\frac{3}{2}}\left(\frac{4n}{e}\right)^n \tag{60.2}$$

It is interesting to compare (51), (53) and (60). We now note that

 $' = 2$   $+ 2$   $+$  (61)

which, using (12.1), (14.1), (24.1), (51.1) and (60.1) may be integrated to give an expression in x. Use of (12.1), (14.1), (29.1), (42.1), (48.1) and (51.1) then tells us that

 $=$  $+ \big(1 +$  $\big) \ln$  $+ \ln^2$ 

$=$  $+$  $\ln$  $+ \ln^2$ 

$=$  $+ \ln$  $+ \ln^2$  (62.1)


$$^{L^n} = 4^n\frac{n!}{2} + \frac{n^2+n-1}{n}\frac{(2n-1)!}{n!} + \frac{n+1}{n}\frac{(2n)!}{n!}\sum_{r=n+1}^{2n}\frac{1}{r} \quad n \geq 1 \sim \sqrt{\frac{\pi n}{2}}\left(\frac{4n}{e}\right)^n \tag{62.2}$$

While we're at it,

$\Big($  $^2\Big)^{L^n}$
$$= \frac{(2n-1)}{n}\frac{(2n-1)!}{n!} + \frac{1}{n}\frac{(2n)!}{n!}\sum_{r=n+1}^{2n}\frac{1}{r} \quad n \geq 1 \quad \sim \frac{\sqrt{2}(1+\ln 2)}{n}\left(\frac{4n}{e}\right)^n \tag{63}$$

Suppose we identify two points on a sphere. By our 'counting' argument, together with (12.1) and (16.1) we obtain an expression which, using (13.1) may be written as

 $= 1 + \ln$  $+ \ln^2$  (64.1)

From (48.1)


$$^{L^n} = 4^n\frac{(n-1)!}{2} + \frac{n-1}{n}\frac{(2n-1)!}{n!} + 2\frac{(2n-1)!}{n!}\sum_{r=n+1}^{2n}\frac{1}{r} \quad n \geq 1 \quad \sim \sqrt{\frac{\pi}{2n}}\left(\frac{4n}{e}\right)^n \tag{64.2}$$

If our object is a finite graph G (with or without loops and multiple edges) containing the point p, we may assume wlog that p is a vertex of G. $G_p(x)$ will then be a polynomial in x with degree at most $\lfloor\frac{1}{2}\deg p\rfloor$, where loops are taken to have degree two. For example, if $p = v$ then

$$L^m = (1+x)^m \quad m \geq 0 \tag{65.1}$$

and

$$(L^m)^{L^n} = \binom{m}{n}n! \quad m, n \geq 0 \tag{65.2}$$

In (65.2) L^n denotes, as usual, the single vertex graph, and *not* $L^n(x)$.

A Slightly More Formal Approach

The rate at which we are accumulating equations may give the impression that we are making rapid progress. Actually, even the number of results that we have *stated* is small. This is because most of our "results" depend on the notions of "drawing" and "sliding", which we have not defined. A natural way to define a drawing is:

A drawing of L^n on T_p is a continuous injective function $d : L^n \to T_p$.

Under this definition, a drawing of L on $\mathbf{R}^2$ is a simple closed curve (a Jordan curve). We could, if we wanted, add other conditions so as to get rid of 'nasty' possibilities such as snowflakes and blancmanges, but for the time being let's assume that no extra conditions have been added. At least two definitions of "sliding" suggest themselves.

(a) Two drawings d_1 and d_2 are homotopic if and only if there exists a homotopy $f : L^n \times I \to T_p$ rel $\{v\}$ from d_1 to d_2 (modulo some edges being turned around), such that for every $t \in I$, f_t is a drawing.

(b) Two drawings d_1 and d_2 are isotopic if and only if there exists an isotopy $g : T_p \times I \to T$ rel $\{p\}$ such that g_0 is the identity, and $d_2 = g_1 d_1$ (modulo some edges being turned around).

In either case, we could leave out the rel condition, which would correspond to allowing the vertex to wander around the object during the sliding. Alternatively, we could replace it by the weaker condition that the image of v has to lie in some fixed set A with $\{p\} \subset A \subseteq T$ during the sliding. When $A = T$, this amounts to the same thing. For general A, it is equivalent to saying that v is always mapped to p, but during the sliding, p is allowed to wander around A provided it ends up back where it started so that nobody notices the difference, which is closer to how we thought of sliding in (32). Note that the possibilities are beginning to mount up. I conjecture that:

(i) In the homotopy case rel $\{v\}$, (12), (14), (16), (20), (23), (29), (37), (42), (45), (46), (47), (51), (53), (62) and (64) hold. Note that we can formalize the turning-over idea.

(ii) In the isotopy case rel $\{p\}$, the same list of equations hold provided we define our objects in a sensible way. By this I mean that (to return to an earlier point)  and  are taken so as not to include their boundary (except for p) and similarly for the other ambiguous objects. We insist upon this so that the edges of L^n can't get 'caught up' on the boundary. Note that yet another way of looking at things would be to place restrictions on where the edges are allowed to go.

(iii) In the weakened homotopy case, (34), (38) and (40) hold for A as suggested in the text.

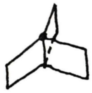
Basically I'm conjecturing that everything goes through modulo the boundary problem. Note that the generating function associated with the object on the left of (35), for example, has not been defined under any of the formulations mentioned. In general if the homotopy definition gives us a generating function, then so will the isotopy definition (put $f_t = g_t d_1$), but intuitively these will usually be quite different. I conjecture that

 $= (x + \sqrt{1-4x})$  **homotopy rel** $\{v\}$ **(66)**

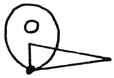
 $= (x+1)$  isotopy rel $\{p\}$ (67)

where the object on the left is obtained from  and a closed unit interval by identifying p with one endpoint and a point q somewhere in the middle with the other endpoint. Intuitively, if the edge pq is not used, then in the former case all edges of L^n can 'pass through' q, whereas in the latter case none can. (66) and (67) were obtained from (12.1) and (13.1). Note that even the seemingly innocent statement that under the isotopy definition (with or without rel), (16.2) holds for $n = 1$ trivially implies Schönflies' theorem for Jordan curves - hence the large number of conjectures! Given that the two definitions are not equivalent, this raises the question of which (if either!) is the best to take. I don't know the answer to this question. When writing this article I could have picked one of them at random and used that, but this seems like a rather silly way of doing things.

Any function which has as its range an indiscrete space is continuous. This observation tells us that if T is an indiscrete space having at least $2^{\aleph_0}$ elements then, under any of the formulations mentioned, $T_p(x) = e^x$. I conjecture that this is also the case if $T = \mathbf{R}^m$ where $m \geq 4$ and that for $m = 3$ the homotopy definition gives us e^x whilst under the isotopy definition $(\mathbf{R}^m)^{L^n}$ is infinite for all $n \geq 1$. That $(\mathbf{R}^m)^L$ is infinite for $m = 3$ under the isotopy definition (without the rel condition) is well known ('there are infinitely many knots') as is the fact that it is one for $m \geq 4$ in the same setup ('knots do not exist in dimension higher than three'). The other cases are probably well known also, but I'll conjecture them anyway. In particular I'm claiming that $(\mathbf{R}^3)^L$ is one under the homotopy definition (with or without the weakened rel condition). The reader who finds this unintuitive should observe that pulling a knot tight is a perfectly good homotopy! Suppose now that we stick three rectangles together along a side and let p be one of the endpoints of this side. Suppose further that for the definition of sliding we stick in the rel condition and that, in order to get a generating function, we take the homotopy definition. The reader may like to try and convince him/herself that

 $= e^x$ (68)

The really ambitious reader may then like to try and convince somebody else of this, which is considerably harder! A counter-example to the assertion that if T_p^L is finite then so is $T_p^{L^n}$ for $n \geq 2$ is shown below for the homotopy definition. Assume that the 'rel $\{v\}$' has been included in the definition, although the counter-example is still valid if we weaken this as mentioned.



For $n > 1$ a 'synthetic hole' may be created by sending the first edge round the triangle. At the time of writing, I do not know if the result is true for the isotopy definition. The reverse implication is certainly false for both definitions, as can be seen by taking T to be a graph with vertices p, q_1 and q_2, and having r edges pq_1, s edges pq_2 and infinitely many edges q_1q_2. For $n = 1$ there are infinitely many possibilities, for $n = 2$ there are $2\binom{r}{2}\binom{s}{2}$ and for $n \geq 3$ there are none at all.

So far we've considered only single vertex graphs and have not mentioned single vertex digraphs. A **digraph** is a graph with directed edges and since there are two ways of directing each edge, this effectively means replacing x with $2x$ everywhere, remembering to multiply by two every time we differentiate anything, and multiplying the right hand side of every equation of the form '$T_p^{L^n} = \ldots$' (where, except in (23) and (65.2), a picture of the object replaces the T_p) by 2^n. Likewise if the '=' is a '$\sim$'. There are some noticeable exceptions, however. (1) and (3) stay as they are, although the meaning of the $T_p^{L^n}$ is different. The right hand side of (6) is halved. (8), (9), (10), (21), (22), (23), (32), (33) and (68) stay as they are. We *do* have

 $'\; =$  ^{2} (69)

(in place of (17) and (18)), so the x in (20.1) disappears and the modified (20.2) is valid for $n \geq 1$. Likewise, the calculation of the generating function associated with the projective plane needs only the observation that

 $'\; = 2$  $+$  (70)

Performing the integration yields (47.1) without the third term. The third term of the modified (47.2) disappears also. (45) and (46) are no longer valid. Finally, (50) and (52) stay as they are except that the terms under the summation signs are multiplied by 2^r. Thus in general we can expect to get nicer results by considering single vertex digraphs, but will have lots of twos floating around, which is why I've written about single vertex graphs.

It is often possible to make guesses and then prove them. For example, we might guess that

$\Big[$  $^m \Big]^{L^n} = \dfrac{m(2n+m-1)!}{(n+m)!} \qquad m \geq 1, n \geq 0$ (71)

after having examined what we get for small m. (71) is equivalent to

 $^m = \displaystyle\sum_{n=0}^{\infty} \frac{m}{2n+m} \binom{2n+m}{n} x^n \qquad m \geq 1$ (72)

which can be shown to be correct by forming a differential equation with boundary conditions for the right hand side, and showing that the left hand side satisfies it. Replacing x by x^2, multiplying by x^m, differentiating, dividing by x^{m-1} and replacing x^2 by x yields

 $^m = \displaystyle\sum_{n=0}^{\infty} \binom{2n+m}{n} x^n \qquad m \geq 0$ (73.1)

$\Big[$  $^m \Big]^{L^n} = \dfrac{(2n+m)!}{(n+m)!} \qquad m, n \geq 0$ (73.2)

There is a saying 'Be wise, generalize'. In this article we've been very general at the expense of being rigorous. Is it possible to generalize further? One thing we could do would be to consider ordered triples (T, p_1, p_2) where T is a (wlog connected) topological space and p_1, p_2 are elements of T. We could define $T^{L^n,L^m}_{p_1,p_2}$ to be the number of ways of drawing L^n at p_1 and L^m at p_2 in such a way that 'intersections' are only allowed at p_1 and p_2. We could also allow one or both of p_1 and p_2 to wander around the object. In the cases for which this is finite for all n and m we could define

$$T_{p_1,p_2}(x_1,x_2) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} T^{L^n,L^m}_{p_1,p_2} \frac{x_1^n x_2^m}{n!m!} \tag{74}$$

This would be a generalization of (1) in the sense that

$$T_{p_1,p_2}(x,0) = T_{p_1}(x) \tag{75.1}$$

$$T_{p_1,p_2}(0,x) = T_{p_2}(x) \tag{75.2}$$

and

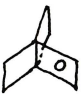
$$T_{p,p}(x_1,x_2) = T_p(x_1+x_2) \tag{75.3}$$

(54), (55) and (56) have already hinted at this idea, and if modified slightly provide expressions for three such generating functions. Another sense in which we can generalize is by allowing L to be an arbitrary space containing v.

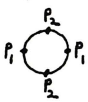
The reader may be under the impression that generating functions are just a tool for getting answers to combinatorial problems. This would be sad. The key word here is 'just'. To me, the much more exciting possibility is that of obtaining properties of the *functions* and relations between them. For example, of the equations above, my favourite is (64.1). Unfortunately this equation involves only 'elementary' functions. How nice it would be if one could apply the techniques used in this article to obtain new equations between functions that aren't elementary, and then prove them rigorously by other means. Oh yes - if anyone wants to send me results not mentioned in this article, they will be gratefully received.

Exercises

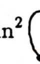
1). Derive (12.1) by showing that  $= 2x$  $'+1$, using (4) and (8), eliminating  $'$ from the resulting equation, and (11) to get a cubic in .
2). Verify (64.1) using (15), (31) and (51.1).
3). Investigate  and  without the rel condition.
4). Investigate , , ,  and the complete graph on m vertices. The first may be helpful for the second.
5). Show that

 $= \begin{cases} \frac{1}{1-x}e^x & \text{graph} \\ \frac{1}{1-2x}e^x & \text{digraph.} \end{cases}$

6). Show that

 $(x,x) =$   ^{4}.

7). Show that

 $= \left(\frac{1+\sqrt{1-4x}}{2\sqrt{1-4x}} + \ln\right.$  $\left.\right)\cdot\left(\frac{1+4x+\sqrt{1-4x}}{2(1-4x)} + \frac{1+\sqrt{1-4x}}{\sqrt{1-4x}}\ln\right.$  $+\ln^2$  $\left.\right)$.

Solutions to Problems Drive

0. G and R swapped. G* mortal, R* devil, B* saint.

1. n=23 (You don't need a calculator - for example $21^{22} < 32^2 \times 21^{20} = (21\sqrt{2})^{20} < 30^{20}$, as $\sqrt{2} < 10/7$. $22^{23} = 8 \times 1331 > 10^4$ so $22^{23} > 10^{92/3} > 2 \times 10^{30} > 2 \times 30^{20}$. Then $21^{22^{23}}$ is *much* bigger than $40^{30^{20}}$.)

2. $n = 5$. 0A: 130130, 3A: 21, 4A: 224, 6A: 302, 8A: 44, 10A: 120440; 0D: 102, 1D: 12, 2D: 3100, 5D: 2142, 7D: 210, 9D: 40.

3. N=3. Any starting position with two pieces in column `B` will do, e.g. `A0`, `B0`, `B1`. (Try it and see!)

4. 4: BDIHGFCE, CDAHFIEG, EHDFBIGA, GBCIAFED.
(Trivially there are $\binom{9}{2} = 36$ pairs, and 8 are used per meeting so there are at most four.)

5. A says "B is a frog. C is a toad"; B says "A is a frog., C is the prince".
(B's strategy is to pretend that C is the prince thus:

A says	*B pretends*	
"B frog, C prince"	A toad, B frog, C prince	In these cases B is pretending to lie, so has 4 choices of statement
"B toad, C toad"	A frog, B frog, C prince	
"B toad, C prince"	A toad, B toad, C prince	B's statement is unique, but the Princess cannot tell A is a frog
"B frog, C toad"	A frog, B toad, C prince	That's it

6. Georg Cantor; Emmy Noether; Leonhard Euler; Sir Isaac Newton; Andrey Nikolayevich Kolmogorov; Baron Augustin-Louis Cauchy; René Descartes; Bernhard Riemann; Karl Weierstrass; Carl Friedrich Gauss; George Green; Évariste Galois; Pierre de Fermat.

7. `QARCH` = 82567, `EUREKA` = 495402.
(A possible approach: let the columns be 0 (units) to 5. From column 5, E is between 1 and 5. Thus for each of the 10 possible values of R, determine in turn the carry from column 2 to column 3, the value of A (which is even from column 0) and the value of H. This gives a small number of easy cases.)

8.

$0 \le n \le 5$	1
$n = 6$	3/4
$7 \le n \le 8$	5/8
$9 \le n \le 10$	1/2
$11 \le n \le 13$	3/8
$14 \le n \le 20$	1/4
$21 \le n \le 40$	1/8
$n > 40$	0

9. A: Quintics; B: Adams; C: New Pythagoreans; D: **Tensors**; E: TMS.

10. (0) 22, 26; (1) 160, 8I0; (2) 2, 2; (3) 30, 12.

((0) Numbers which are a product of distinct primes; (1) Factorials in base 24; (2) Number of primes in factorisation; (3) Dates of Easter from 1989 to 1998†)

11. Icosahedron: $OQS > 90$; Octahedron: $OQS = 90$; Tetrahedron: $OQS < 90$. (For the octahedron, let P,Q,R be the centres of 3 faces of a cube meeting at a vertex S...)

This year's winner was Frazer Jarvis, who was (in fairness to the other competitors) handicapped by Graham Nelson. They will set next year's questions.

Solution to Chess Puzzle

N	P↑	P↑			P↑	P↑	N
P↓	P↓					P↓	P↓
				R			
B		K					B
B					K		B
			R				
P↑	P↑					P↑	P↑
N	P↓	P↓			P↓	P↓	N

The above shows that the answer is at least 28. Suppose we can get 29 pieces onto the board. Then only 3 pieces are left, and so there are at least 3 rooks (or queens, which are worse) on the board. Together, they rule out 3 rows and 3 columns, which is $6 \times 8 - 9 = 39$ squares. This leaves 25 squares for the remaining 26 pieces, which is impossible.

There really is very little spare space on that board. I think the only other solutions are the ones you get by swapping a knight with the adjacent pawn in the same row.

Solution to A, B, C, D Problem

$A = 2, B = 5, C = 9, D = 2$. The author has not yet proved uniqueness except by use of computing.

† Oddly enough, no-one got this one - Ed.

The Cover

"Hausdorff Dale"
by Tim Auckland

The cover is a montage of the following fractals. Note that D is the fractal dimension of the object in question.

The Earth ($D \sim 1.2$)	The zero-set of a familiar $S^2 \to \mathbf{R}$ fractional Brown function
Feignbaum Cascade ($D = 0...1$)	A plot of the non-wandering set of the real quadratic map against the free parameter, showing period doubling transition to chaos
I.F.S. Trees (D not globally defined)	The attractor of a random sequence of applications of four contractive affine maps
I.F.S. Plume ($0.63 < D < 1$)	As above, but two maps now have overlapping images, producing a Cantor set
Recursive Trees (No D)	Trees drawn by recursively defining the trunk shape
Peano Dragon ($D = 2$)	0L-System with axiom "X" and production rules $X \to X + YF+, Y \to -FX - Y$ where $F=$ draw line, $+=$ turn right by $\frac{\pi}{2}$, $-=$ turn left by $\frac{\pi}{2}$
Bird in a Thorn Bush ($D = 1+$)	(X, Y) picture of the attractor of the dynamical system $X \to (1-c)\cos(\pi a Y) + cZ$, $Y \to X$ with $a = 1.99$, $Z \to Y$ and $c = 0.8$
Lorenz Attractor ($D = 2+$)	$(X - Y, Z)$ picture of the attractor of the differential system $X' = 10(Y - X)$, $Y' = 28X - Y - XZ$, $Z' = XY - \frac{8Z}{3}$
Mandelbrot Clouds (D not known)	Affine transformations of the set of $c \in \mathbf{C}$ s.t. the Julia set of $z \to z^2 + c$ is connected
Puppy ($D = 1$)	A small dog with a beard

It was calculated and assembled on an Archimedes 440, and then transferred via ten different formats or so across various different computers before printing.