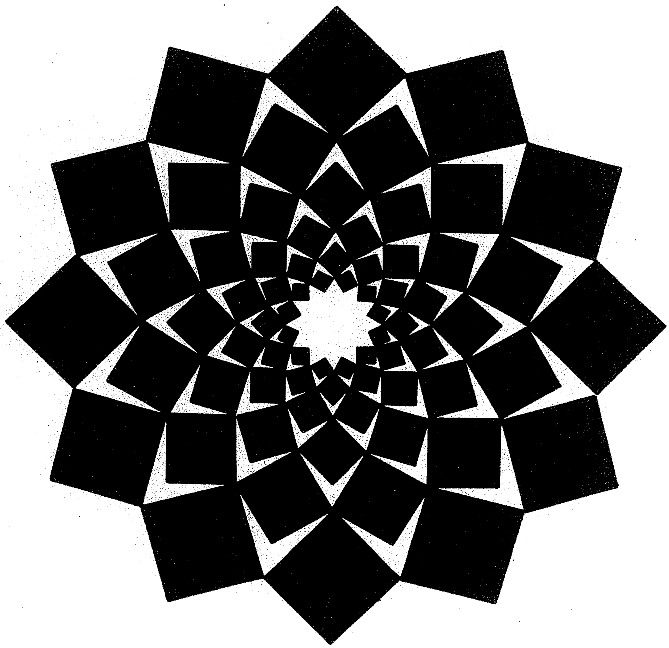
ISSN 0313-6825

FUNCTION

Volume 2 Part 3

June 1978



A SCHOOL MATHEMATICS MAGAZINE

Published by Monash University

Function is a mathematics magazine addressed principally to students in the upper forms of schools. Today mathematics is used in most of the sciences, physical, biological and social, in business management, in engineering. There are few human endeavours, from weather prediction to siting of traffic lights, that do not involve mathematics. *Function* contains articles describing some of these uses of mathematics. It also has articles, for entertainment and instruction, about mathematics and its history. Each issue contains problems and solutions are invited.

It is hoped that the student readers of *Function* will contribute material for publication. Articles, ideas, cartoons, comments, criticisms, advice are earnestly sought. Please send to the editors your views about what can be done to make *Function* more interesting for you.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

EDITORS: G.B. Preston (chairman), N.S. Barnett, N. Cameron, M.A.B. Deakin, B.J. Milne, J.O. Murphy, G.A. Watterson, (all at Monash University); N.H. Williams (University of Queensland); D.A. Holton (University of Melbourne); E.A. Sonenberg (R.A.A.F. Academy); K.McR. Evans (Scotch College, Melbourne)

BUSINESS MANAGER: Joan Williams (Tel. No. (03) 541 0811, Ext. 2591)

ART WORK: Jean Hoyle

Articles, correspondence, problems (with or without solutions) and other material for publication are invited. Address them to:

The Editors,
Function,
Department of Mathematics,
Monash University,
Clayton, Victoria. 3168

Alternatively correspondence may be addressed individually to any of the editors at the addresses shown above.

The magazine will be published five times a year in February, April, June, August, October. Price for five issues (including postage): $3.50; single issues 90 cents. Payments should be sent to the business manager at the above address: cheques and money orders should be made payable to Monash University. Enquiries about advertising should be directed to the business manager.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

Registered for posting as a periodical - "Category B"

The leading article in this issue, "The Winds over the Earth", is written by one who has been closely involved in and has made significant contributions to the solutions, which are not yet final, to problems of explaining the general pattern of air-flow in the earth's atmosphere. Professor Priestley, a Fellow of the Royal Society and of the Australian Academy of Science, is a leading meteorologist, from 1949 until recently Chairman of the (Australian) National Committee on Meteorology and, in 1967, Chairman of the World Meteorological Organisation.

In *Function* we often deal with mathematics which is clearly outside the high school syllabus and we have several times heard the comment from teachers "That's too difficult. That cannot be understood by High School kids". In most cases this comment has been made because the subject of the article appeared to be about mathematics outside the school syllabus. The editors know the school syllabus and take special pains to ensure that in fact all our articles are readable by our intended audience. Just because an article is about or uses mathematics not in the present school syllabus does not mean that it cannot be explained in an intelligible and interesting fashion. We are trying to bring you articles of interest about mathematics. Do not be put off if at first sight the material seems unfamiliar.

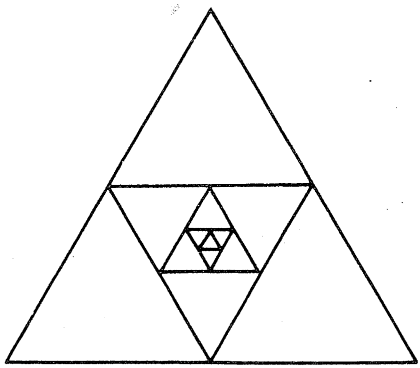
(See the inside back cover.)

CONTENTS

THE FRONT COVER

J.O. Murphy, Monash University

An illustration of geometric progression is given by the diagram on the front cover where successive groups of twelve squares of the same size are arranged on the circumference of a series of concentric circles. One mode of construction is first to draw the outer circle which is then divided into twenty-four parts with alternate points of subdivision forming two opposite vertices of a square and the associated chord representing the diagonal. This diagonal is bisected by the radius drawn to the intermediate point of subdivision and measuring this distance inwards and outwards from the centre locates the other two vertices and also the radius of the next inner circle for the next group of twelve squares. The division of an equivalent triangle, as illustrated on this page, also demonstrates a geometric progression for the relationship between the sides and also the areas of the triangles.



The common vertices of successive squares of different sizes lie along what is called a logarithmic or equiangular spiral. The usual polar form of the equation to this spiral is given by

$$r = ae^{\theta \cot\alpha},$$

where α is the constant angle (hence the name equiangular) the curve makes with the radius vector - i.e. the angle between the radius vector and tangent.

Four flies sit at the corners of a square table, with sides of length ℓ, facing inward. They all start walking at the same time, at the same rate, and each directing its motion toward the fly on its right. Taking the centre of the table as a suitable origin use a geometric argument to show that the flies all move along equiangular spirals. (The actual path solutions $r = \ell e^{-(\theta+\frac{\pi}{4})\sqrt{2}}$ etc. can be determined from the differential equation for the path.)

THE WINDS OVER THE EARTH

C.H.B. Priestley, Monash University

In an article in *Function* (August 1977, Volume 1, Part 4), Chris Fandry began to describe how mathematics is used in the understanding and prediction of atmospheric winds and ocean currents. In particular he derived the approximate relation between the wind and the pressure, namely that the wind blows with speed proportional to the horizontal pressure gradient, with high pressure to the left* and low pressure to the right* of its path.

The following illustration might help you to understand this law. Imagine a ruler held steady above a record player in motion and used to draw a line slowly on the record from the centre outwards. An observer in 'fixed space' will know that this is a straight line (corresponding to motion under no forces), but the record will show a line curving progressively to the left (i.e. anticlockwise when the record is viewed from above). To have kept the pencil in a straight line *on the record* would have required a force acting from left to right. The pressure gradient provides such a force for a parcel of air moving steadily relative to our rotating earth.

The relation is not merely one in which the pressure field determines the winds. As Fandry described, the pressure is *hydrostatic* and equal to the weight of air above the position in question: higher (or lower) pressure therefore means that more (or less) air has accumulated above. So the relation between the pressure and the movement of the air is one of those chicken-and-egg relationships in which one cannot uniquely identify one as the cause and the other as the effect.

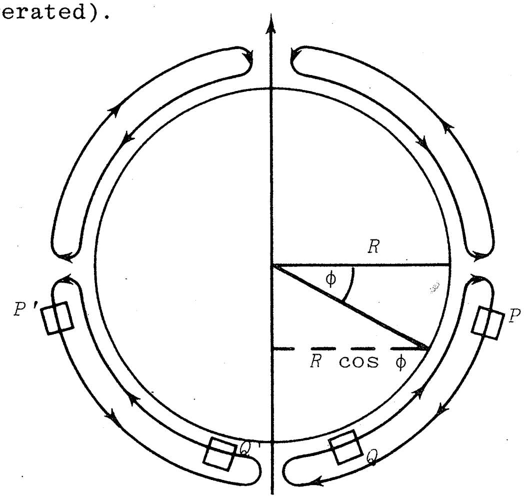
We shall here use this principle and two others to help put together the quite complicated pattern of prevailing winds and pressure distribution over most of the face of the earth. The atmosphere rotates, to first approximation, with the rotating earth. R, the radius of the earth, is $6{\cdot}37 \times 10^6$ metres and Ω, the angular velocity, is $7{\cdot}29 \times 10^{-5}$ radians per second. At the equator, then, the circumferential speed (ΩR) is 464 metres per second. The velocity of the wind, or *relative* motion of the atmosphere, is much smaller than this, generally by almost two orders of magnitude.

If the surface of the globe was uniform, and absorbed radiation from the sun everywhere at an equal rate, there would be no relative motion, no wind. But on average the

---

*We word this article for the southern hemisphere: for the northern, replace 'left' by 'right' and vice versa throughout Left and right here are relative to a person travelling along the path.

solar radiation at the equator is greater than at the poles making the surface of the earth warmer at the equator than at the poles. We now introduce our second principle, namely that in a fluid which is unevenly heated from below a circulation is set up, rising where it is warmer and hence lighter and sinking where it is colder (heavier), with flow from the colder to the warmer near the bottom and returning at the top. This would appear on the earth as a circulation in meridional planes in the pattern indicated in the diagram (the relative depth of the atmosphere is greatly exaggerated).



But this idea of a purely meridional circulation has to be modified to take rotation into account, through considerations of angular momentum. The angular momentum of a mass M, rotating with absolute circumferential velocity v at a distance r from the axis of rotation, is Mvr. Our third principle is that the laws of motion, as applied to rotating bodies, state that the rate of change of angular momentum is equal to the *torque* or 'twisting moment'. The torque is defined as the component of force in the circumferential direction multiplied by the distance from the axis of rotation, r. A good example to illustrate torque is the use of a wrench to tighten a bolt. In the absence of torque, angular momentum will be conserved. A good example of such conservation is that of an ice skater twirling on the point of one skate, throwing out arms and the other leg to slow up, rotating faster when she brings these into one line to decrease r, and so increase v.

In the atmospheric case, let u be zonal (i.e. east or west, west positive) component of wind speed relative to the earth's surface and v the actual speed, and we have, in latitude ϕ,

$$r = R \cos \phi$$
$$v = u + R\Omega \cos \phi$$

and we have seen that $R\Omega$ = 464 metres per second. We take, in our model, the average circulation as zonally symmetric and will consider the whole atmosphere as divided into zonal rings, two of which are indicated in the diagram as PP' and QQ'. Note that the pressure can exert no net force, and hence no net torque, on a complete ring because one notional circuit of the ring returns us to the same point and the same pressure.

Now consider what would happen to the complete ring PP' moving polewards as in the diagram. We suppose it to start from the equator with zero zonal wind. On reaching latitude ϕ, conservation of angular momentum requires that

$$(u + R\Omega \cos \phi)R \cos \phi = R^2\Omega$$

whence $$u = \frac{R\Omega \sin^2 \phi}{\cos \phi}.$$

You may verify that this is equal to 32, 134, 328 metres per second at latitudes 15°, 30°, 45° respectively.

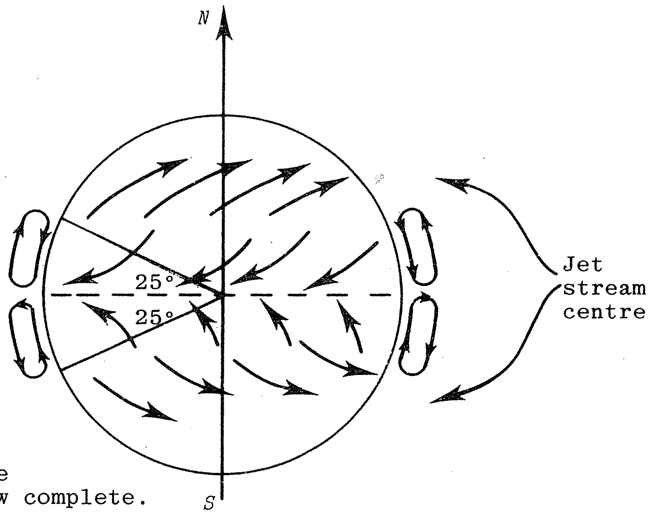
A corresponding surface ring such as QQ', moving equatorward, would develop an easterly wind (negative u). However in this case there would simultaneously develop a resisting torque due to friction against the earth's surface, and the speed of the surface easterlies would be more restrained.

Our picture is now that of a single cell in each hemisphere, producing steady SE surface winds (NE in the northern hemisphere), westerly upper winds increasing with latitude, and persistent ascent in the equatorial region accounting for the very high rainfall there. This is known as the Hadley cell, after the first man to explain it; the surface winds are called the *trade* winds. For tropical latitudes this is a good approximation to the actually occurring state. However the Hadley cell is unable to extend into middle latitudes beyond about 25° for two reasons which we shall next describe.

As the ring PP' approaches the mid-latitudes the upper winds do indeed become very strong westerlies, as calculated. But it can be shown that this *jet stream*, as it is called, is dynamically unstable due to the immense shearing motions, and that it must break up into large horizontal wave-like disturbances. As in the break-up of any other fluid flow the disturbances propagate and, the atmosphere being relatively shallow like the skin of a fruit, the disturbances quickly take over the whole depth of atmosphere in the latitudes where they occur.

The second reason for breakdown is that, in a single cell extending to the poles, the zonal surface wind would be easterly in all latitudes. Therefore the only torque (surface friction) applied to the whole atmosphere would be uniform in direction, altering the total angular momentum of the atmosphere progressively in one sense; while there would be an equal and opposite torque on the earth, continuously slowing its rotation rate. Such would conflict with the requirements of a balanced and steady average state. For a steady state, given that there is a zone of surface easterlies there must also be a zone of surface westerlies, supplying an equal and opposite torque.

For these two reasons, then, there is generated in middle latitudes a highly disturbed flow which is subject to one basic constraint: there must on average be a prevailing westerly wind. The mechanism for providing it lies in the fact that the disturbances have strong vertical motions within them. The annular rings are broken up, as it were, and there is a much greater mixing between the different layers of atmosphere which brings downwards a substantial part of the angular momentum from the very strong prevailing westerlies at upper levels.



The essential model of our earth's wind systems, or general circulation of the atmosphere, is now complete. The Hadley cell extends to about 25° on average, with the two trade winds converging in the equatorial zone where the rising motion brings heavy rainfall. The trade winds are steady compared with the disturbed middle latitudes which have very strong upper westerlies and, *on average*, a westerly surface component. Between the low and middle latitude regimes lies the descending branch of the Hadley cell. High surface pressure must be built up here to provide pressure-wind balance as discussed at the beginning. Descending air and high (i.e. anticyclonic) pressure are characteristic of clear dry weather. This is why most of the desert regions of the world are to be found near latitudes 25 to 30° North and South. As the sun moves with season it strives to draw the circulation with it, so that each feature of the pattern is found some 5 to 10° further south in January and further north in July.

Many of the zones bear names from the days of sailing ships trading between Europe, India, and Cathay: the *trades* themselves, the *doldrums* in between, where calms were too common, the *horse latitudes* (30°) where becalming was even more frequent and prolonged, water could grow short and horses were thrown overboard. In the southern hemisphere westerlies, stronger than the northern westerlies because they had more sea fetch and less mountain shelter, were called the *roaring forties*. These winds often could not be faced on the return from Australia, and had to be avoided by zigzags far to the north or by continuing eastwards round perilous but wind-helpful Cape Horn. The routing of ships and aircraft according to wind still remains one of the functions of the meteorological services.

It was in 1735 that Hadley published his theory, which remains today the accepted one for almost half of the earth's surface (can you prove that half the area lies between ±30°?). For the other half, many 'half-baked' theories were offered in the ensuing period, but it took 215 years before a reasonably complete and acceptable scientific explanation, along the lines indicated here, was advanced. So any difficulty the reader has had sets him in good company historically. It must also be realised that the jet-stream, an essential part of the full picture, was not discovered until the mid 1940's when radiosondes and a few high flying aircraft began to penetrate to heights around 10 kilometres. The mathematical proof of the instability of its long waves followed only a few years later.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

ARITHMETICKE-RITHMETICALL, OR THE HAND MAID'S SONG OF NUMBERS

G.C. Smith, Monash University

Text books writers in all times have been troubled that their readers may not find their work as interesting or memorable as it should be. Various ways of improving the interest and appearance of text books have been tried: in recent years there have been notable improvements in the standard of text figures, and the use of colour. One device I have not seen in recent books is the use of verse; but in older works it was tried from time to time.

One of the earliest pieces of arithmetic in verse is said to be from a manuscript of Elizabethan times (Elizabeth I not II!), about the year 1570 in fact:

> Multiplication is mie vexation,
> And Division is quite as bad,
> The Golden Rule is mie stumbling stule,
> And Practice drives me mad.

The golden rule is perhaps better known as the *rule of three*, or the Merchant's rule: it has a long history, for it can be found in the work of the Indian Brahmagupta who lived in the 7th century. In modern terms it amounts to solving the proportion $x:a = b:c$ for x given a,b,c - these are the 'three' occurring in the name of the rule. 'Practice' is a term used in English texts which describes problems which can be solved by the rule of three.

I do not know whether this verse ever appeared in a printed book, but the next examples occur in Thomas Hylles (pronounced Hills, I expect) 'The Arte of Vulgar Arithmeticke, both in integers and fractions...' published in London in 1600. Like many books of this time the title goes on and on for several lines. Whenever Hylles states a rule or theorem, he breaks into verse:

> All primes together have no common measure
> Exceeding an ace which is all their treasure.

'An ace' here means 1: the rule is that prime numbers have no common divisor but 1.

He explains addition and subtraction of fractions as follows:

> Addition of fractions and likewise subtraction
> Requireth that first they all have like basses
> Which by reduction is brought to perfection
> And being once done as ought in like cases,
> Then adde or subtract their tops and no more
> Subscribing the basse made common before.

Here 'basse' means denominator, 'top' means numerator.

Nicolas Hunt wrote 'The Hand-Maid of Arithmetic refined...' which came out in London in 1633. His verses are headed by the words of the title of this article. A sample of his 'arithmeticke-rithmeticall':

> Subtract the lesses from the great, noting the rest
> Or ten to borrow, you are ever prest,
> To pay what borrowed was thinke it no paine,
> But honesty redounding to your gaine.

This, I think, needs no further explanation.

The next item comes from Peter Halliman's 'The Square and Cube Root compleated and made easie'. The second edition appeared in 1688. The object of this book is to find square and cube roots by tables, which were derived from the approximate formula:

$$\sqrt[n]{(a^n + b)} \simeq a + \frac{b}{(a + 1)^n - a^n}$$

(where $n = 2$ or 3). The curious aspect of it is that the author seems to think this formula gives *exact* square and cube roots! His comment on his achievement he gives in verse:

Now Logarithms lowre your sail
And Algebra give place,
For here is found, that ne'er doth fail,
A nearer way, to your disgrace.

At a later period some authors posed problems in verse form. Charles Vyse's 'The Tutor's Guide, being a complete system of Arithmetic' was a popular text. It was first published in 1799 and went to at least eleven editions. The following verse problem is said to be the best known of its day:

When first the Marriage-Knot was tied
Between my Wife and me,
My Age did her's as far exceed
As three Times three does three;
But when ten Years, and Half ten Years,
We Man and Wife had been,
Her Age came up as near to mine
As eight is to sixteen.
Now, tell me, I pray
What were our Ages on the Wedding Day?

The solution? I leave that to you.

Would any reader of *Function* care to try their hand at putting the chain-rule or De Morgan's laws into verse? I am sure the Editor would be interested in any ventures into the arithmeticke-rithmeticall.

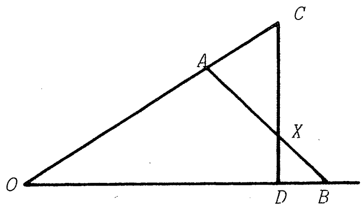
∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

THE EMBLEM

J.C. Barton, University of Melbourne

In the early 1930's there was a University Lecturer named M.L. ("Mac") Urquhart, who gave lectures in Mixed Mathematics at the University of Melbourne. He was a likeable personality, very dear to his students, and his discourses on hexagonal cities, sex, electromagnetism and relativity were regarded very highly by those bright young scholars who were fortunate enough to sit at his feet. This they did literally for it was customary for Urquhart to enter the lecture room, seat himself on top of the lecture table in front of the class, then begin to talk about the world while rolling a cigarette, until such time as he considered that it might be in order to continue the previous session's lecture on div, grad and curl, dipoles and those other mathematical things which so comfortably explained the world of physics in those uncomplicated days of the thirties. He was interested in the theory of relativity.

From this arose his interest in proving a minor theorem, namely that if $OA + AX = OD + DX$, then $OC + CX = OB + BX$. It had something to do with the times of travel of light



rays between O and X. The diagram was subsequently elevated into an emblem of the Mathematical Association of Tasmania, of which body Mac Urquhart was sometime President.

Urquhart said that the theorem was the easiest thing to prove. It's not, though neither is it inordinately difficult. The reader may care to try to prove it before reading further.

At this stage let Dr Fred Syer take up the story. He writes as follows. Historically, Mathematical Associations in Australia are copies of The Mathematical Association (of Great Britain) which, founded in 1871, was the successor to the Association for the Improvement of Geometrical Teaching. When "Euclid" was a school subject the statement that if $OA + AX = OD + DX$ then $OC + CX = OB + BX$ would not have needed elaboration, but with the subsequent improvements (which have nearly wiped the subject out) one should add that (OAC), (ODB), (AXB), (CXD), that is, each triad of points in parentheses, are collinear, that OA etc. are signless magnitudes (or positive real numbers) denoting measures of length of the line-intervals between the named end-points. It is to be noted further that the diagram can just as readily illustrate the theorem that if $OA - AX = OD - DX$, then $OX - CX = OB - BX$. We suppose in the following constructions and proof that A and D are on opposite sides of OX. Should we choose A and D on the same side of OX, the statements of the theorems read thus:

$$\text{if } OA + AX = OD + DX \text{ then } OC - CX = \pm(OB - BX);$$

alternatively,

$$\text{if } OA - AX = OD - DX \text{ then } OC + CX = OB + BX.$$

CONSTRUCTIONS AND PROOF

[For the following proof the reader needs to know what is meant by an escribed circle of a triangle. Such a circle touches the three lines determined by the three sides of the triangle, but lies entirely outside the circle. Any triangle thus has three escribed circles: draw an example of this.

You also need to use the fact that the two tangents that can be drawn from any point outside a circle to the circle are of equal length.]

It would spoil the austere beauty of the emblem (writes Dr Syer) to do other than describe the constructions (each to conclude with a triumphant Quod Erat Faciendum) before the consequent Quod Erat Demonstrandum. The reader is urged to supply his own constructions.

Let a circle escribed to triangle OAB touch OA in P, OB in Q, AB in R.

Let the tangent through X [other than the line AB] to this circle touch it in S and meet OA in C', OB in D'. Then

$$\begin{aligned} OA + AX &= OP - AP + AR - XR \\ &= OQ - XS \\ &= OD' + D'Q - (D'S - D'X) \\ &= OD' + D'X. \end{aligned}$$

But $OA + AX = OD + DX$, as given.

Hence $OD + DX = OD' + D'X$, or

$$D'D = D'X - DX.$$

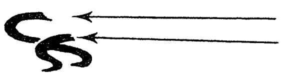
That is, in the triangle $D'DX$ one side is equal to the difference between the other two. It follows that D and D' coincide. Now will

$$\begin{aligned} OC + CX &= OP - CP + CS + SX \\ &= OQ + RX \\ &= OB + BQ + BX - BR \\ &= OB + BX, \text{ as required.} \end{aligned}$$

Supposing, instead of the above, that the circle inscribed in triangle OAB were to touch OA in P', OB in Q', AB in R' and that the tangent through X were to meet this circle in S' and to meet OA in C'', OB in D'', it could be similarly shown that C'' coincides with C and D'' with D so that $OA - AX = OD - DX$ and that consequently $OC - CX = OB - BX$, as forecast.

This completes Dr Syer's very elegant demonstration of the theorem.

SOLUTIONS TO PROBLEM 1.3.2



forward right wheel
back right wheel

Submitted by Christian Cameron,
Year 9, Glen Waverley High School.

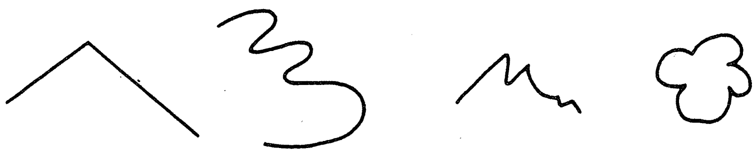
CONTINUOUS CURVES

G.B. Preston

We are all familiar with continuous curves. The path of a cricket ball or an aeroplane, or of a bird or a fly, or again the shape of a length of string, are all continuous curves. Surely we all know what a continuous curve is. In two dimensions it is the sort of curve we can draw on a piece of paper without taking the pencil off the paper.

Can a precise definition be given of what a continuous curve is? Can we capture the intuitive idea lying behind the examples of continuous curves I have just given? A definition is given at school today which originated in the early nineteenth century. Baron Cauchy, a French mathematician, defined, in the 1820's, a continuous function in the way we do today. Let us give his definition and use it to define a continuous curve.

A function F is said to be *continuous at* b if there is no jump in its value at b, in other words, if $F(t)$ can be made as close to $F(b)$ as we please provided t is close enough to b. Examples of graphs of functions that are not continuous at b are shown in Figure 1.



In example (i) there is a jump in the value of $F(t)$ as soon as t becomes greater than b. In example (ii) the value $F(b)$ of F at b is right off the curve represented by the other values of $F(t)$ for t not equal to b. Again F has a jump in its values at b. We are tacitly assuming in our discussion that F is defined over an interval of numbers that includes b.

To have F continuous at b we have to exclude any jump, however small, in the values of F, as t increases from just less than b to just greater than b. Thus, to show that a function F is continuous at b it will suffice to show that, for any (and every) possible jump, if t is close enough to b then $F(t)$ differs from $F(b)$ by an amount less than the jump being considered; for then there certainly is not a jump in the values of F at b as big as the jump being considered.

---

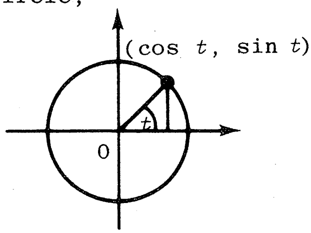
† Part of a talk given to fifth and sixth formers at Monash University, 29th April, 1977.

The function F is said to be *continuous* if it is continuous for all b for which $F(b)$ is defined.

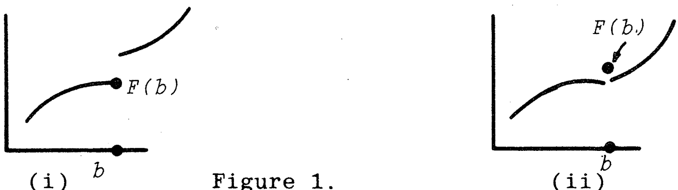
We shall make use of this definition in constructing some interesting continuous curves.

The graph of a continuous function is a special case of a continuous curve. The graph of a function F is the set or *locus* of all points $(t, F(t))$, for the relevant values of t. Let us generalize this and look at two functions E and F, say, and consider the locus of all points $(E(t), F(t))$. If E and F are each continuous functions of t for some interval of values of t, say, for $0 \leqslant t \leqslant 1$, then the locus of all points $(E(t), F(t))$ for $0 \leqslant t \leqslant 1$, is what we define to be a *continuous curve*. In this we can replace $0 \leqslant t \leqslant 1$ by any other interval of values of t. The idea can be extended to curves in three dimensions.

As an example consider the locus of (cos t, sin t) where $0 \leqslant t \leqslant 2\pi$. The locus is then a circle, centre the origin and radius 1 (since $\cos^2 t + \sin^2 t = 1$). Note that we really have extended the idea of a graph of a continuous function in our definition of a continuous curve. For this circle is certainly not the graph of all points $(t, F(t))$ for any function F - because there are two points on the circle corresponding to each value, with two exceptions, of the first coordinate.



Has Cauchy's definition captured just what we had in mind when we began trying to define a continuous curve? Well this depends on what we wanted. We have arranged that our curves have no jumps in them - the continuity of E and F ensures this. But is this all we want? Our definition allows curves like the following,



(i) Figure 1. (ii)

smoothly turning curves, or pieces of such curves put together with junctions at which there are no tangents. For a long time it was thought that this was essentially the kind of curve that had been defined.

In 1856, Karl Weierstrass, a German mathematician, made one of the most famous discoveries in the history of mathematics: he discovered a continuous curve which had a tangent at no point. Such a curve is impossible to imagine visually. In 1890, Giuseppe Peano, an Italian mathematician, made what was perhaps an even more startling discovery: he constructed a continuous curve which not merely had no tangent at any point, but also passed through every point inside a square. Thus continuous curves in the plane are not even one-dimensional, they can fill a square.

Indeed, Peano's construction quickly extends to give curves in three-dimensions which pass through every point inside a cube, *space-filling* curves.

In fact it is easy to construct a pathological curve with the bizarre properties of Peano's curve and, with the wisdom of hindsight, it is hard to believe that it took mathematicians so long to realise what they were talking about. Why were they so shocked to discover space-filling curves and curves with a tangent at no point? At least two reasons suggest themselves. First, they were not at all clear what they were trying to do in defining continuous curves; in other words, they had not clarified at all just what idea they were trying to capture when a precise definition was suggested for such curves. Second, the idea of a function was still mixed up with the idea of a formula for calculating its values. Functions were thought of as things like $x^2 + 1$, $\sin 3x$ etc., and combinations of them. The idea that a function, in the definitions that were actually used, was just a special kind of correspondence, was slow to emerge. Weierstrass's continuous curve without a tangent at any point was an infinite sum of sines and cosines, what is called a Fourier series: $a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \ldots + a_n \cos nx + b_n \sin nx + \ldots$, in which the coefficients $a_0, a_1, b_1, a_2, b_2, \ldots, a_n, b_n, \ldots$ were chosen so as to give a curve with the required properties. The feeling that a function had to be defined by some kind of formula delayed an understanding of what continuous functions, in Cauchy's sense, are.

Weierstrass's and Peano's constructions seem to show that Cauchy's definition, although it captured part of the intuitive idea of a continuous curve, had really missed essential properties that still needed to be pinned down. For example, a "curve" surely just has to be one-dimensional. The story of how mathematicians got a little nearer to this elusive idea of a curve is an interesting one, but I am not going to tell it now. Instead I shall end my comments by justifying my claim that a space-filling continuous curve with a tangent at no point is easy to construct.

We shall construct such a curve in the plane. The curve we construct will consist of all points $(E(t), F(t))$, for $0 \leqslant t \leqslant 1$, where E and F are continuous functions (yet to be defined) which will have the property that for any point (a,b) with $0 \leqslant a \leqslant 1$ and $0 \leqslant b \leqslant 1$, there is a t such that $E(t) = a$ and $F(t) = b$. Thus every point (a,b) in the unit square lies on this continuous curve.

We now define E and F; the construction we give is due to Peano. The interested reader may like to consult the original paper, entitled, "Sur une courbe, qui remplit toute une aire plane" (On a curve which completely fills a plane area) which appears in 1890 the mathematical journal *Mathematische Annalen*, Volume 36, pages 157 - 160. The paper is beautifully clear and includes an expanded version of what follows here.

We are to define E and F for all values t in the closed interval $0 \leqslant t \leqslant 1$. Each number t must first be expressed in base 3:

$$t = 0\cdot t_1 t_2 t_3 \ldots, \tag{1}$$

where each t_i is one of the digits 0, 1 or 2 and where we write (1), in analogy with the decimal notation (where the base is 10), to mean that

$$t = \frac{t_1}{3} + \frac{t_2}{3^2} + \frac{t_3}{3^3} + \ldots \quad . \tag{2}$$

For example, if $t = 1/3$, then $t = 0\cdot10000\ldots$; again if $t = 1/4$, then $t = 0\cdot020202\ldots$.

We shall call such an expression (1), with the meaning given by (2), a *trimal*, and say that t is then expressed as a trimal.

Then when $t = 0\cdot t_1 t_2 t_3 \ldots$, a trimal, we define $E(t)$ and $F(t)$, thus:

$$E(t) = 0\cdot e_1 e_2 e_3 \ldots,$$

$$F(t) = 0\cdot f_1 f_2 f_3 \ldots,$$

where $e_1 = t_1$ and, for $n = 2,3,\ldots,$

$$e_n = \begin{cases} t_{2n-1}, & \text{if } t_2+t_4+\ldots+t_{2n-2} \text{ is even,} \\ 2-t_{2n-1}, & \text{if } t_2+t_4+\ldots+t_{2n-2} \text{ is odd,} \end{cases}$$

and, for $n = 1,2,3,\ldots,$

$$f_n = \begin{cases} t_{2n} & \text{if } t_1+t_3+\ldots+t_{2n-1} \text{ is even,} \\ 2-t_{2n}, & \text{if } t_1+t_3+\ldots+t_{2n-1} \text{ is odd.} \end{cases}$$

As examples, consider $t = 1/4$ and $t = 1/5$. Since, as a trimal,

$$1/4 = 0\cdot020202\ldots,$$

the above definitions give

$$E(1/4) = 0\cdot000\ldots,$$

$$F(1/4) = 0\cdot222\ldots .$$

And, since, as a trimal,

$$1/5 = 0\cdot012101210121\ldots,$$

we have

$$E(1/5) = 0\cdot0000\ldots,$$

$$F(1/5) = 0\cdot1111\ldots .$$

We have glossed over a small but important point in the above definition. Some numbers have two different expressions as trimals. For example

$$0\cdot0022000\ldots = 0\cdot0021222\ldots,$$

or, again,

$$0\cdot211000\ldots. = 0\cdot210222\ldots .$$

In fact this is the only way in which a number can be expressed as a trimal in two ways: any trimal which ends in an infinite sequence of 0's is equal to another trimal ending in an infinite sequence of 2's. The digit 2, for trimals, is playing the same role as the digit 9, for decimals. Note, however, that $t = 1 = 0\cdot222\ldots$, only has one expression in the required form.

Let us now show that when any number t, such that $0 \leqslant t \leqslant 1$, is expressed as a trimal in two different ways, then the rules we have just given for constructing $E(t)$ and $F(t)$ give the same values whichever expression of t as a trimal is used. We give the argument in smaller print and it may be skipped by a reader prepared to accept the result we are proving.

So, let t have two expressions as a trimal. Thus, we can assume that,

$$t = 0\cdot t_1 t_2 \ldots t_{2n-2} t_{2n-1} t_{2n} 222\ldots, \quad (3)$$

and

$$t = 0\cdot t_1 t_2 \ldots t_{2n-2} t'_{2n-1} t'_{2n} 000\ldots, \quad (4)$$

where not both t_{2n-1} and t_{2n} are equal to 2. Let us consider the value of $E(t)$. There are two cases to consider.

(1) $t_{2n} \neq 2$; when it follows that $t_{2n-1} = t'_{2n-1}$ and $t'_{2n} = t_{2n} + 1$; (2) $t_{2n} = 2$; when, since t_{2n-1} is then not equal to 2, $t'_{2n} = 0$ and $t'_{2n-1} = t_{2n-1} + 1$.

(1) $t_{2n} \neq 2$. Since $t_{2n-1} = t'_{2n-1}$, the first n places of the construction of the trimal for $E(t)$ are the same for either expression of t as a trimal. Put $t_2+t_4+\ldots+t_{2n} = a$. Then, if a is even, the digit in the $(n+1)$-th place in the expression for $E(t)$, as constructed from (3), is 2; and since $a + 2 + 2 + \ldots + 2$ is then always even, every successive digit after this is also 2.

But then, for the construction of $E(t)$ from (4), $t_2 + t_4 + \ldots + t_{2n-2} + t'_{2n} = a + 1$ is odd, and so the digit in the $(n+1)$-th place is then $2 - 0 = 2$; and since $(a + 1) + 0 + 0 + \ldots + 0$ is then always odd, every successive digit after this is also 2.

Thus, starting from either expression (3) or expression (4), we get precisely the same $E(t)$. The argument is similar, and leads to the same conclusion, when a is odd.

(2) $t_{2n} = 2$. In this case we get two distinct expressions for $E(t)$, but they are two trimals of equal value.

Since the two trimals for t are identical for the first $2n - 3$ places the two constructions for $E(t)$ agree for the first $n - 1$ digits. Put $t_2 + t_4 + \ldots + t_{2n-2} = b$. Then, if b is even, the digit in the n-th place in the expression for $E(t)$, as constructed from (3), is t_{2n-1}, and in all further places it equals 2. Similarly, the n-th digit in the expression of $E(t)$, as constructed from (4) is t'_{2n-1}, and all further digits are then equal to 0. Since $t'_{2n-1} = t_{2n-1} + 1$, the two expressions for $E(t)$ have the same value. If b is odd, then, beginning with the digit in the n-th place, the construction of $E(t)$ from (3) gives $(2 - t_{2n-1})000\ldots$, while the construction from (4) gives $(2 - t'_{2n-1})222\ldots$. Since $t'_{2n-1} = t_{2n-1} + 1$, so that $2 - t_{2n-1} = (2 - t'_{2n-1}) + 1$, these two expressions for $E(t)$ are equal trimals.

This completes the argument for E; the argument for F is similar.

Let us call our curve C: thus C is the locus consisting of all points $(E(t), F(t))$ for $0 \leqslant t \leqslant 1$. We now show that C passes through every point in the unit square. If (a,b) is any such point, and as trimals, $a = 0{\cdot}a_1a_2\ldots$, and $b = 0{\cdot}b_1b_2\ldots$, say, then it may be checked that the following construction gives a number t with $E(t) = a$ and $F(t) = b$. Take $t_1 = a_1$, and then

$$t_2 = \begin{cases} b_1, & \text{if } a_1 \text{ is even,} \\ 2 - b_1, & \text{if } a_1 \text{ is odd,} \end{cases}$$

$$t_3 = \begin{cases} a_2, & \text{if } b_1 \text{ is even,} \\ 2 - a_2, & \text{if } b_1 \text{ is odd,} \end{cases}$$

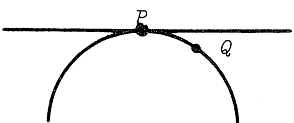
and generally,

$$t_{2n-1} = \begin{cases} a_n, & \text{if } b_1 + b_2 + \ldots + b_{n-1} \text{ is even,} \\ 2 - a_n, & \text{if } b_1 + b_2 + \ldots + b_{n-1} \text{ is odd,} \end{cases}$$

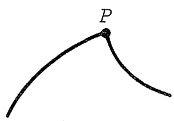
$$t_{2n} = \begin{cases} b_n, & \text{if } a_1 + a_2 + \ldots + a_n \text{ is even,} \\ 2 - b_n, & \text{if } a_1 + a_2 + \ldots + a_n \text{ is odd.} \end{cases}$$

We can now show that E and F are continuous for all t in $0 \leqslant t \leqslant 1$. For suppose that $0 \leqslant t \leqslant 1$ and $0 \leqslant u \leqslant 1$, and that both u and t are in trimal form. Consider t to be fixed. Then, as u approaches t, the trimal form for u approaches the trimal form (or one of the trimal forms if there is more than one) of t; in other words the first n digits of u coincide with the first n digits of t, expressed in (some) trimal form, and as u tends to t, n tends to infinity. Thus as u tends to t so $E(u)$ tends to $E(t)$ and $F(u)$ tends to $F(t)$. In other words E and F have no jump in their values at t; and this argument holds for all t in $0 \leqslant t \leqslant 1$. Thus E and F are continuous.

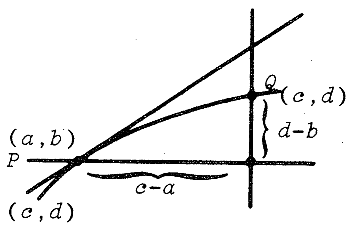
A curve is said to have a *tangent* at a point P if it can be approximated by a single straight line for all points on the curve sufficiently close to P. If P is the point (a,b) and we take a point Q on the curve, with coordinates (c,d), say, close to P, then Q must be, if there is a tangent to the curve at P, almost on the tangent line at P (see accompanying picture). In other words (again see picture) $d-b$, for varying (c,d) close to (a,b), must be almost proportional to $c-a$ (it would have to be exactly proportional for Q on the tangent line).



Tangent at P.



No tangent at P.



Now, consider our curve C, at a point $P = (a,b) = (E(t), F(t))$, say. Take a neighbouring point $Q = (c,d) = (E(u), F(u))$ with u close to t. We can take u close to t, i.e. Q close to P, by taking the first digits, say the first $2n$ digits, in the trimal expansion of u, to be the same as the first $2n$ digits in the trimal expansion of the given number t, determining P. Thereafter u can have any entries in its trimal expansion. By taking n larger and larger we get a point Q closer and closer to P. After the $2n$th place, the digits in the odd positions of the expansion of u have no connexion with the digits in the even positions. So we can choose the digits in u, from the $(2n + 1)$th place on, so that $d-b = F(u) - F(t)$ has a value which has no connexion with that of $c-a = E(u) - E(t)$ (and where we can assume, for the purpose of the calculation of $d-b$ and $c-a$, that $E(u)$, $E(t)$, $F(u)$, and $F(t)$ each have their first n digits equal to 0).

Hence, as (c,d) approaches (a,b), i.e. for Q close to P, the value of $d-b$ is certainly not almost proportional to the value of $c-a$. Hence there is no tangent at $P = (E(t), F(t))$. This argument holds for all points P on the curve.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

REVIEW OF "TOOLS FOR THOUGHT" BY C.H. WADDINGTON

DAVID DOWE, GEELONG GRAMMAR SCHOOL

At the present time our understanding of the way the world reacts to our best efforts to change it (if possible, for the better) seems at its least ever. We lend considerable sums of money to an under-developed country and show it how to organize public health, and the result is that the level of nutrition falls alarmingly and, due to an increase in population without an increase in available food, babies, instead of dying of infectious diseases as they did before, now die of starvation.

Tools for Thought asks us to stop looking at the world as a set of cause-and-effect relationships, but rather as a series of interacting systems. We too often think that what goes on around us can be understood as a set of simple causal sequences where A causes B, B causes C, C causes D, etc. But this will suffice only when A causes B but has little effect on anything else, and similarly when the overwhelmingly most important effect of B is to cause C. In the example cited above, we see that when modern health care is applied to a primitive society, its effect is so powerful that it is no longer only a question of curing a few cases of illness; the population increases rapidly, due to a decreased death rate of the young, and thus demands more food.

Waddington does not claim to be able to solve the world's problems, nor to know what they are, but instead offers methods of looking at the problems, methods which he terms 'tools for thought'. Perhaps the most important 'tool for thought' that Waddington describes is philosophy. Two differing views, the 'thing' view - that the world consists essentially of 'things', and that any changes we notice are secondary, and arise from the way they (the things) interact with one another - and the 'process' view - that the things we discern are only stills out of what is essentially a movie - are given, and Waddington argues that neither view is always 'correct' and that we should use whichever one is appropriate to the situation. The chemical atoms of iron, carbon, oxygen, etc. will not change their essential nature in any period of time we are normally concerned with; so it is simpler to regard them as things (not processes), and get down to the practical problem of how they interact with one another to bring about the secondary processes of chemical reaction. On the other hand, when looking at living systems, the 'process' view is to be preferred - we are more interested in the fact that muscles can contract to bring about bodily movement than we are in the chemistry by which the contraction is produced.

A new idea, concerning complexity of information in the modern world, is the *coefficient of immediacy* - the ratio of the increase in a variable to its value at the end of that period. For instance, if the amount of information doubles in 15 years, it would be A at the beginning and $2A$ at the end of that period. Thus, the coefficient of immediacy is $\frac{A}{2A} = \frac{1}{2}$. Thus, if someone's education finished at the beginning of that 15-year period, he would be 50%-obsolescent at the end of the period (unless he continued learning in the meantime). Over a working life of 45 years, a person in this situation would become 87·5% obsolescent.

There are many other tools for thought which Waddington offers - feed-back in systems, information theory, handling systems (games theory), and a great many others - but I thought it better to go into detail on one or two than attempt to cover them all equally. The hope of the author is that one day man will be able to solve what Waddington calls the 'World Problematique' - the entire world problem - by sufficiently extending our tools for thought.

$5.60 and in Paladin paperbacks, *Tools for Thought* is both relevant and necessary to the world of today. I advise you to buy a copy.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

Twenty-four car makers and importers are each fighting for a dwindling share of the market.

This Day Tonight, ABC, May 15, 1978.

DISMAL SCIENCE, DEMOGRAPHIC DESCRIPTIONS AND DOOMSDAY

M.A.B. Deakin, Monash University

In 1798, Thomas Malthus first proposed his theory of population growth. His work is now seen as the first theoretical advance in the science of *demography* (the study of population trends), as well as being an important early contribution to economics. Essentially, Malthus saw population as increasing according to a compound interest (or exponential, or geometrical) law, while resources remain fixed. That is, the earth's resources are finite, so that there exists an upper bound to the population the earth can support.

Malthus' account of population growth is the simplest available in the sense that every other plausible one constitutes an adjustment to his basic law. We may follow that account in mathematical language rather than use the verbal descriptions on which he himself relied. Let $N = N(t)$ be the world population at some time t in the future. Although, strictly speaking, $N(t)$ can only take on integer values, and is thus discontinuous (i.e. altering in jumps rather than varying smoothly), we may, to an excellent approximation, replace it by a continuous "trend curve", which it is our aim to analyse.

Suppose that in each year every 1000 people experience (for example) 40 births and 30 deaths. At the end of the year there are 1010 people for every 1000 at the beginning. More generally, in a short time interval Δt, N people will experience $BN\ \Delta t$ births and $DN\ \Delta t$ deaths, so that the change in population, ΔN, is given by

$$\Delta N = (B - D)N\ \Delta t, \qquad (1)$$

or, letting $\Delta t \to 0$,

$$\frac{dN}{dt} = rN\ , \qquad (2)$$

where $r = B - D$. In honour of Malthus, r is termed the *Malthusian parameter*.

Equation (2) is a differential equation, i.e. the function $N(t)$ is the unknown, to be determined from information about its derivative. The solution of equation (2) is (check by differentiating)

$$N = N_0 e^{rt}, \tag{3}$$

where N_0 is some constant. N_0 is in fact the value of N at the time $t = 0$. (When precisely t does equal 0 is a matter of our own convenience - we can start our clock whenever it best suits us to do so. N_0 is the value of N at the instant the clock begins to run.)

For $r > 0$ (the case we are interested in), a graph of N against t gives the steeply rising curve of Figure 1. It is, of course, not feasible for the population to continue to rise in this way. Indeed, it cannot do so beyond some value K, the maximum carrying capacity of the earth. The earth, on this model would become saturated at time T, satisfying

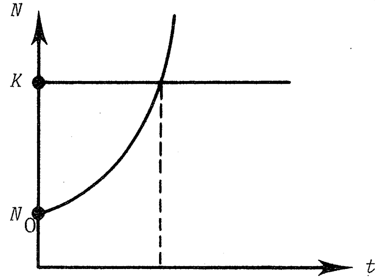


Figure 1

$$N_0 e^{rT} = K . \tag{4}$$

When $t > T$, $N = K$. The population is held at this level because any additional people die for want of sustenance.

This pessimistic conclusion led to economics becoming known as "the dismal science". Nevertheless, it remains an important part of our thinking today, and a worrying problem, now that world population has got much closer to K. (The actual value to be assigned to K is the subject of considerable debate.) Because the same analysis applies to animal and plant populations, whose numbers in the undisturbed wild are often relatively constant, it follows that individual plants and animals compete for the available spaces (or *niches*, as they are called).

It was precisely this thought that prompted Charles Darwin to formulate his principle of natural selection - the survival of the fittest, one of the keys to our modern account of evolution. Interestingly enough, this theory was put forward independently by Alfred Russell Wallace, and he too was influenced by Malthus.

This is a very successful record of application, yet Malthus' law (3) cannot possibly be true. Suppose that in 10,000 B.C. the human population had consisted of a single couple (it was in fact larger than this, of course) and had continued to grow at a rate of 1% per annum. If this process had continued unchecked, the earth would now be a solid sphere of human flesh engulfing the whole solar system and much of interstellar space besides - in fact the solar system would be a mere speck at its centre. (Can you duplicate this calculation?)

The present rate of increase in world population is about 2%; Malthus assumed the even higher figure of $2\frac{3}{4}$%

(approximately - he had a doubling time, the time taken for N to grow from N_0 to $2N_0$, of 25 years). 1% would appear, on this showing, to be a very conservative estimate, but clearly it is far too high. In the past, for most of the time, the percentage of annual increase must have been much lower than this. It follows that it has varied over time, and from this it follows that the Malthusian parameter r has not remained constant, but has varied also, so that equation (3) is not the solution of equation (2).

Were we discussing a problem in Physics, such a discrepancy between theory and fact would be regarded as very damaging and the theory would be discarded. In the present context, however, the matter is not so serious. We may well reply that Malthus is describing a tendency, a useful one to know about, but of course other factors intervene - wars, plagues, famines and what not deplete the population, while inventions like agriculture, flush toilets and antibiotics tend to allow it to increase. We may well say that it is unreasonable to expect the same value of r to apply to paleolithic man, to Malthus' contemporaries of the industrial revolution and to ourselves.

This enables us to conclude that, *as a description of actual events*, Malthus' law is limited in its applicability to short time spans without major changes or upheavals. Despite this, it is useful and valid *as an account of a natural tendency* over long periods of time and even for many species other than man.

There have, however, been attempts to improve on Malthus, and most of these try to better the record of equation (3) as a description. One influential modification was developed independently by P.F. Verhulst in 1844 and Raymond Pearl in 1924. One way to arrive at the Pearl-Verhulst law (as it is called) is to modify equation (2). Both of these investigators proposed that dN/dt, as well as being proportional to N, was also proportional to the fraction of niches still available for habitation. They therefore replaced equation (2) by

$$\frac{dN}{dt} = rN\left(\frac{K - N}{K}\right). \qquad (5)$$

This equation has the solution (check by differentiation)

$$N = \frac{KN_0}{N_0 + (K - N_0)e^{rt}}. \qquad (6)$$

The graph of this function is shown in Figure 2 for the practical case $N_0 < K$. Pearl, in particular, attempted to show that this curve was an improvement on equation (3) as a description of a wide range of population data.

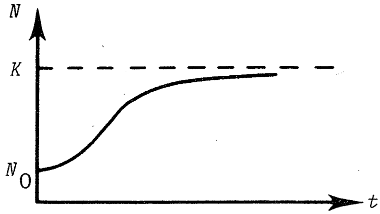


Figure 2

By and large, this endeavour has been unsuccessful. The Pearl-Verhulst law (6) (now better known as the logistic equation) has a certain place in theoretical ecology, and fruit flies kept under strict laboratory conditions can be induced to follow it, but it has not been markedly more successful than Malthus' law (3) for most human populations. Its best success was an extremely accurate (within 1%) prediction of the population of the U.S.A. in 1940 made by Verhulst almost 100 years earlier. Its worst failure was the determination by Pearl and Reed of a value for K of 2×10^9; the world population is now over double this figure.

There have been a number of other amendments, the most startling of which appeared in 1960. This was the work of three electrical engineers: Heinz von Foerster, Patricia Mora and Lawrence Amiot. According to their view, mankind is continually altering the natural world to suit its own expansion, so that N should be increasing *faster* than allowed for by Malthus.

To allow for this, they wrote

$$\frac{dN}{dt} = rN^{1+1/k}, \tag{7}$$

where k was a positive constant.

Equation (7) has the solution (again check by differentiation)

$$N = \frac{A}{(t_0 - t)^k}, \tag{8}$$

where $A = (\frac{k}{r})^k$ and t_0 is a constant. By using statistical methods of curve fitting, the three authors used data on world population, mostly over the period since 1800, to estimate A, k and t_0.

The value of t_0 they found was A.D. 2026.87, i.e. the 14th of November 2026. That date should be doomsday, for as a glance at equation (8) or Figure 3 will show, on that day the world population becomes infinite! In view of this conclusion, von Foerster and his coworkers felt justified in moving the date forward to Friday, the 13th of November. (It really is a Friday.)

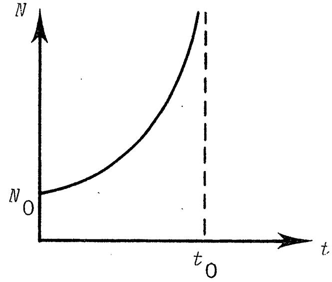


Figure 3

Many commentators assumed that the article was a hoax, but this seems not to have been the case. Indeed, the world population is increasing faster than the Malthusian law (3) allows. We can test this by plotting log N against t. If Malthus is correct, the result should be a straight line; in point of fact the points lie on a curve that bends up above the straight.

The doomsday model gives a very good fit indeed. Although in 1962, it was still widely regarded as due for a swift disproof, it was checked in both 1970 and 1975 and found to give very accurate results. If anything, the world population was slightly ahead of schedule for doomsday.

Of course, the same comment applies here as to the absurd Malthusian prediction. The model's usefulness as a predictor extends only to limited time spans without major disturbing factors. Whether the doomsday model will turn out to have the theoretical importance of Malthus' analysis is highly doubtful.

Certainly, we may be assured that the world's population will not be infinite in 2026; even so, the question of overpopulation allows us little room for complacency.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

PROBLEM 3.1

As a classroom project, two students keep a calendar of the weather, according to the following scheme: Days on which the weather is good are marked with the sign +, while days on which the weather is bad are marked with the sign -. The first student makes three observations daily, one in the morning, one in the afternoon and one in the evening. If it rains at the time of any of these observations, he writes -, but otherwise he writes +. The second student makes observations at the same times as the first student, writing + if the weather is fair at any of these times and - otherwise. Thus it would seem that the weather on any given day might be described as ++, +-, -+ or -- (the first symbol made by the first student, the second symbol by the second student). Are these four cases all actually possible?

AN INTERESTING TRICK

T.M. Mills

At this year's Easter fair in Bendigo, my son won a "magic calculator" in a side-show. Naturally I confiscated it until I figured out how it worked.

The calculator consists of six cards which are shown below. Show all the cards to a friend and ask him or her to select one number from any card. If your friend points out those cards on which the selected number appears, then you can determine the number by adding together the numbers in the top left hand corners of those specified cards.

Why does the trick work?

Devise a similar scheme for the numbers from 1 to 15 inclusive.

1	3	5	7	9	11	13	15
17	19	21	23	25	27	29	31
33	35	37	39	41	43	45	47
49	51	53	55	57	59	61	63

2	3	6	7	10	11	14	15
18	19	22	23	26	27	30	31
34	35	38	39	42	43	46	47
50	51	54	55	58	59	62	63

4	5	6	7	12	13	14	15
20	21	22	23	28	29	30	31
36	37	38	39	44	45	46	47
52	53	54	55	60	61	62	63

8	9	10	11	12	13	14	15
24	25	26	27	28	29	30	31
40	41	42	43	44	45	46	47
56	57	58	59	60	61	62	63

16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63

32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

NUMBER THEORY

Geoffrey Chappell
Kepnoch High School

Numbers are the most essential elements of mathematics; without them, not much of mathematics would exist. So it is only fitting that numbers be treated in this magazine.

The field of number theory is enormous so in this article I shall deal with a small but important area that has many applications such as the solution of Problem 5.4.

The theory of congruences was developed by Carl Friedrich Gauss in 1801 in "Disquisitiones Arithmeticae". If a and b are integers and m is a natural number, then a and b are congruent mod m if $m|(a - b)$, i.e. m is a factor of $a - b$. This is written as $a \equiv b \pmod{m}$.

Congruences may be combined; for instance if we let $a \equiv b \pmod{m}$, then $ka \equiv kb \pmod{m}$ and $a^n \equiv b^n \pmod{m}$. Also, if c_i is an integer, $c_i a^i \equiv c_i b^i \pmod{m}$ and by adding these together for various i, $c_0 + c_1 a + \ldots + c_n a^n \equiv c_0 + c_1 b + \ldots + c_n b^n \pmod{m}$, or if f is a polynomial function, with integral coefficients,

$$f(a) \equiv f(b) \pmod{m}, \text{ if } a \equiv b \pmod{m}. \quad (1)$$

Equation (1) leads to a relationship between congruences and equations. Consider the polynomial equation,

$$f(x) = 0, \quad (2)$$

where $f(x)$ has integer coefficients, and the polynomial congruence,

$$f(x) \equiv 0 \pmod{m}. \quad (3)$$

The connexion between (2) and (3) is close: if x_0 is an integer solution of (2) then $f(x_0) \equiv 0 \pmod{m}$ for any m and so x_0 is a solution of (3). The integer solutions of $f(x) = 0$ are all to be found among the solutions of $f(x) \equiv 0 \pmod{m}$.

Also if $x \equiv a \pmod{m}$ is the only solution to (3) then the solutions of (2) are congruent to $a \pmod{m}$ and if there are no solutions to $f(x) \equiv 0 \pmod{m}$ there can be no integer solutions to $f(x) = 0$.

Example: $14x^4 + 7x^3 + x^2 - 3 \equiv 0 \pmod 7$ has no integer solutions because $14x^4 \equiv 0 \pmod 7$ and $7x^3 \equiv 0 \pmod 7$ and $x^2 - 3 \equiv 0 \pmod 7$ has no solutions. Therefore there are no integer values of x satisfying $14x^4 + 7x^3 + x^2 - 3 = 0$.

[The reader should check that to find any solutions to $x^2 - 3 \equiv 0 \pmod 7$ it suffices to check merely $x = 0,1,2,3,4,5,6$.]

Consider the linear congruence $ax \equiv b \pmod m$. Solving this for x is equivalent to finding integers x,y that satisfy.

$$ax = b + my \quad \text{or} \quad ax - my = b.$$

Hence, if x,y satisfying these equations exist, then each divisor of a and m divides $ax - my$ and must therefore be a factor of b. If we let (a,m) denote the greatest common divisor of a and m, then $(a,m)|b$. This may be shown to be a necessary and sufficient condition for $ax \equiv b \pmod m$ to have a solution x.

Example: $2x \equiv 3 \pmod 6$ is not soluble since $(2,6) = 2$ and 2 is not a factor of 3.

It is not as easy to find a condition for solubility of quadratic congruences such as

$$Ax^2 + Bx + C \equiv 0 \pmod p.$$

Multiplying by $4A$ (completing the square), this gives

$$(2Ax + B)^2 \equiv B^2 - 4AC \pmod p.$$

This suggests solving a general form: for primes p, find x such that

$$x^2 \equiv a \pmod p. \qquad (4)$$

If (4) is soluble, a is called a *quadratic residue mod p.* If (4) is not soluble, a is called a *quadratic non-residue mod p.*

Let us introduce what is called Legendre's symbol $(\frac{a}{p})$, defined as follows, for odd primes p:

$$\left(\frac{a}{p}\right) \equiv \begin{cases} 1 \text{ if } a \text{ is a quadratic residue mod } p, \\ -1 \text{ if } a \text{ is a non-residue mod } p, \\ 0 \text{ if } a \equiv 0 \pmod p. \end{cases}$$

There are two important results, concerning Legendre's symbol. The proofs are long (too long for this article) and the reader is advised to see the book "Number Theory" by T.H. Jackson (published by Routledge and Kegan Paul, 1975).

Theorem 1. $(\frac{a}{p}) \equiv a^{\frac{1}{2}(p-1)}$ (mod p), if p is an odd prime.

Theorem 2. $(\frac{p}{q}) = (\frac{q}{p})$ if either p or $q \equiv 1 \pmod 4$

$= -(\frac{q}{p})$ if both p and $q \equiv -1 \pmod 4$ and p and q are different odd primes.

We can use the sign of $(\frac{a}{p})$ to test for solubility of (4). If $(\frac{a}{p}) = 1$, (4) is soluble. It does not indicate what the solution is.

Example: Problem 5.4 (Volume 1).

Since $x^2 - x + 41$ is always odd, then if we assume it to be composite, we can find an odd prime p such that

$$x^2 - x + 41 \equiv 0 \pmod p \qquad (1)$$

or $$(2x - 1)^2 \equiv -163 \pmod p.$$

The Legendre symbol $(\frac{-163}{p})$ must be calculated.

$$(\frac{-163}{p}) \equiv (-163)^{\frac{1}{2}(p-1)} \pmod p \equiv (-1)^{\frac{1}{2}(p-1)}.163^{\frac{1}{2}(p-1)} \pmod p$$

$$\equiv (\frac{-1}{p}).(\frac{163}{p}) \pmod p.$$

Let $p \equiv 1 \pmod 4$

then $(\frac{-1}{p}) = (-1)^{\frac{1}{2}(p-1)} = 1$

and $(\frac{163}{p}) = (\frac{p}{163})$ by Theorem 2

$\equiv p^{81}$ (mod 163) by Theorem 1.

Let $p \equiv -1 \pmod 4$

then $(\frac{-1}{p}) = -1$ and $(\frac{163}{p}) = -(\frac{p}{163})$,

so $(\frac{-1}{p})(\frac{163}{p}) = (\frac{p}{163}) \equiv p^{81} \pmod{163}$.

(1) has a solution iff $p^{81} \equiv 1 \pmod{163}$.

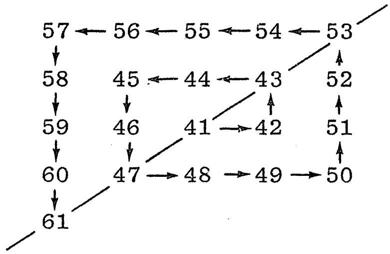
Although there are methods for solving this congruence, it is not difficult to use a calculator to find p^{81} (mod 163) for various primes p.

There is no value of p less then 41 that satisfies (1) i.e. there is no factor of $x^2 - x + 41$, less than 41, this means that for $x^2 - x + 41 < 41^2$, $x^2 - x + 41$ is prime; this occurs when $x < 41$, so:

$$\underline{x^2 - x + 41 \text{ is prime if } x \leqslant 40.}$$

[Note: $x^2 - x + 41$ has an interesting geometrical interpretation. Consider the following spiral:

The diagonal gives values of $x^2 - x + 41$.]



SOLUTION TO PROBLEM 1.4

Glenn Merlo, now at Melbourne University, sent us two programmes, one for a hand calculator, which he had used to verify, as asked in the problem, that for integers up to 1000, certain sequences always end with 4,2,1 repeating. The Fortran program he used was:

```
  DO 3 K=3,999,4
  IF(((K-3)/16)*16.EQ.K-3)GO TO 3
  I=K
1 IF(((I/2)*2).EQ.I)GO TO 2
  I=3*I+1
  GO TO 1
2 I=I/2
  IF(I.LT.K)GO TO 3
  GO TO 1
3 CONTINUE
  PRINT 4
4 FORMAT(/42H 3X+1 THEOREM PROVED FOR 1ST. 1000
                              NUMBERS/)
  STOP
  END
```

SOLUTION TO PROBLEM 2.1 (solution by Geoffrey J. Chappell, Kepnoch High School).

Since we want to find the distance the fly has flown, we are interested in its speed only. Its speed is constant, twice the man's, so in a given time it will have flown twice as far as the man has walked. The man has walked from *A* to *B*:

The fly has flown twice the distance from *A* to *B*.

SOLUTION TO PROBLEM 2.2 (solution by Geoffrey J. Chappell).

Imagine the 3 × 3 × 3 cube to be divided into 27 cubes; it is clear that there must be a cube in the middle that has no exposed surfaces. To form this cube 6 new surfaces must be created and since no two of these surfaces can be formed by one cut, six is the minimum possible number. We know it can be done in 6, so the answer is 6.

Explanatory note: the last sentence is required, because the previous reasoning showed only that the number of cuts $\geqslant 6$. To prove that the minimum is indeed 6, we must know (or show) that it is possible to do it in 6 cuts.

SOLUTION TO PROBLEM 2.3 (solution by Geoffrey J. Chappell).

By inspecting the combinations of wine colours that occur when a knight and his two neighbours are considered, it is seen that:

THEOREM: A knight is left without wine if and only if there is a pair of knights with wine of the same colour.

So to prove that a knight will be left without wine we only have to show that there must be a pair at the table (by pair, I mean a pair of knights with wine of the same colour).

If there are an even number of knights, the wine *could* be arranged alternately around the table and then there would be no pair and hence by the theorem, no knight left without wine.

If there are an odd number of people at the table, there must be a pair and hence a knight left without wine. Assume there are n knights. If the first $n-1$ are not arranged alternately we have a pair already; if they are, the n-th will create a pair. So there must be a pair.

50 knights - there need not be a knight left without wine

51 - there must be.

PROOF OF THEOREM: There are eight combinations; writing them with the considered knight in the middle and his two neighbours on their respective sides, they are:

WWW, *WWR*, *RWW*, *RWR*,
WRW, *WRR*, *RRW*, *RRR*.

By testing each combination, the considered knight is left without wine iff the situations *WWR* and *WRR* exist. (1)

So a knight is left without wine only if there is a pair. If there is a multiplet, then either it is a pair or it contains pairs, and assuming both colours are present, then at the ends of the multiplet we have one of the situations *WWR* or *WRR*, (i.e. *RWW*...*WWR* for a white multiplet or *WRR*...*RRW* for red). By statement (1) a knight is left without wine if there is a pair. Q.E.D.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

PROBLEM 3.2

What point on the earth's surface is farthest from the earth's centre?

PROBLEM 3.3

Blackjack or Twenty-one. Two players in turn take from a pile of 21 matches. At each turn a player must take at most 5 matches and at least 1 match. The player who takes the last match wins.

Devise a winning strategy for playing this game. Generalize.

PROBLEM 3.4

Spot the fallacy:

Since $\cos^2 x = 1 - \sin^2 x$,

it follows that

$$1 + \cos x = 1 + (1 - \sin^2 x)^{\frac{1}{2}},$$

that is $$(1 + \cos x)^2 = \{1 + (1 - \sin^2 x)^{\frac{1}{2}}\}^2.$$

In particular, when $x = \pi$, we have

$$(1 - 1)^2 = \{1 + (1 - 0)^{\frac{1}{2}}\}^2,$$

or $$0 = (1 + 1)^2$$

$$= 4.$$

PROBLEM 3.5

If a side of a triangle is of length less than the average length of the other two sides, show that its opposite angle is less, in magnitude, than the average of the other two angle magnitudes.

∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞ ∞

"It all started back in '66 when they changed to dollars, and overnight me overdraft doubled. I was just gettin' used to this when they brought in kilograms or somethin' and the woolclip dropped by half. Then they started playin' around with the weather and brought in Celcius, and we haven't had a decent fall of rain since. This wasn't enough - they had to change us over to hectares and I end up with less than half the farm I had. So one day I sat down and had a think. I reckoned that with daylight savin' I was workin' eight days a week so I decided to sell out. Then to cap it all off, I'd only just got the place in the agent's hands when they changed to kilometres and I find I'm too flamin' far out of town anyway!"

From *The West Darling Pastoralist* as quoted in the CSIRO Division of Mathematics and Statistics Newsletter, No.42, May 1978.

"...it is essential that he [the teacher] know the questions that do not have a solution, on which he has to be silent. A person who does not know well the foundations of any part whatever of mathematics will always remain hesitant, with an exaggerated fear of rigour."

"Mathematical rigour is very simple. It consists in affirming true statements and in not affirming what we know is not true. It does not consist in affirming every truth possible."

"Of the science which we do know, we have the obligation of teaching only that part which is most useful to our students."

"...even where a definition can be made, it is not always useful to do so. Every definition expresses a truth, namely that an equality can be established whose first member is the entity being defined and whose second, or defining member, is an expression composed of entities already considered. Now, if this truth has a complicated form, or is in some way difficult to explain, it belongs to that order of truths which are not suitable to be taught in a given school, and which may be passed over in silence."

"...any book whatever, even one full of blunders, may be made rigorous by leaving out what is false; what remains is the useful part of the book. We see, then, that rigour produces simplicity and economy in teaching."

"The new methods are difficult, not for the pupils, but for the teachers. They are obliged to make an effort to study them, and to look, from a new point of view, at things which were formerly presented to them under a different aspect. We are all occupied with our own work, and unable to follow up every novelty. Some, on seeing such novelties, can correctly say: 'These are indeed beautiful, but I am unable to study them.' On the other hand, others pretend to have infused knowledge: they say *a priori* that the book is difficult, that they do not understand it, and how then can they make the pupils understand it?"

Giuseppe Peano: *On the foundations of analysis*, 1910.

∞ ∞ ∞ ∞ ∞ ∞