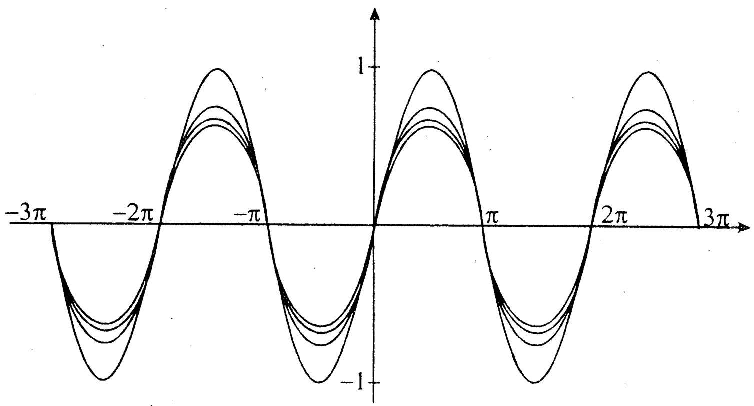
Function

A School Mathematics Journal

Volume 23 Part 1 February 1999



Department of Mathematics & Statistics – Monash University

Reg. by Aust. Post Publ. No. PP338685/0015

Function is a refereed mathematics journal produced by the Department of Mathematics & Statistics at Monash University. The journal was founded in 1977 by Prof G B Preston. *Function* is addressed principally to students in the upper years of secondary schools, and more generally to anyone who is interested in mathematics.

Function deals with mathematics in all its aspects: pure mathematics, statistics, mathematics in computing, applications of mathematics to the natural and social sciences, history of mathematics, mathematical games, careers in mathematics, and mathematics in society. The items that appear in each issue of *Function* include articles on a broad range of mathematical topics, news items on recent mathematical advances, book reviews, problems, letters, anecdotes and cartoons.

* * * * *

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

The Editors, *Function*
Department of Mathematics & Statistics
Monash University
PO BOX 197
Caulfield East VIC 3145, Australia
Fax: +61 3 9903 2227
e-mail: function@maths.monash.edu.au

Function is published five times a year, appearing in February, April, June, August, and October. Price for five issues (including postage): \$20.00* ; single issues \$5.00. Payments should be sent to: The Business Manager, Function, Department of Mathematics & Statistics, Monash University, Clayton VIC 3168, AUSTRALIA; cheques and money orders should be made payable to Monash University.

For more information about *Function* see the journal home page at `http://www.maths.monash.edu.au/~cristina/function.html`.

---

* \$10 for *bona fide* secondary or tertiary students.

EDITORIAL

Welcome to all readers to our new volume of *Function*!

The front cover shows the first few terms of the sequence of the graphs of functions generated by the composition of the function $\sin(x)$ with itself:

$$\sin(x),\ \sin(\sin(x)),\ \sin(\sin(\sin(x))),\ \sin(\sin(\sin(\sin(x)))),\ \ldots$$

We leave to the readers to analyse this sequence. Some features, like periodicity, are obvious. What other features can you see? Can you explain them? Are there any similarities with the sequence generated by the function $\cos(x)$?

There are three feature articles in this issue: an article analysing the best return strategy in a football-tipping competition; an example from optics to illustrate the interaction between science and mathematics; and an article by our young contributor Mark Nolan presenting his research on a sum involving inverse tangents of integers which results in π.

In the regular *History of Mathematics* column you will find some insight into the classical problem of determining the maximum area in a plane enclosed by a string of fixed length. The *Computers and Computing* column gives a simple example to show how parallel processing can be used to speed up lengthy computing tasks.

Finally, we include, as usual, problems and solutions prepared by our recently appointed *Problem Corner* editor John Jeavons. We thank Peter Grossman for the hard work he put in over the last five years to make the *Problem Corner* a success. Also, we take this opportunity for thanking our former editor Bruce Henry for his valuable contributions to so many issues of *Function*.

Happy reading!

* * * * *

BEATING THE ODDS

Michael A B Deakin, Monash University

This article was prompted by the News-Item in *Function Vol* 22, *Part* 2, p. 56 describing a football-tipping competition. I was also reminded of an article run in *Function* many years ago.[1] Much of what I say here derives from that earlier article, but it will be presented in a rather different way.

Suppose we place a bet of \$1 on one team to beat another in a head-to-head contest.[2] Call the teams A and B, say. The bookmaker will reward us with a dividend $\$D_A$ if we successfully back Team A or else a dividend $\$D_B$ if we successfully back Team B. Of course, if we back a losing team, then our dollar is gone!

I took the dividends in the following table from those published by National Sportsbet just before AFL Round 4.

Collingwood	\$2.30	Geelong	\$1.70
Richmond	\$1.55	Western Bulldogs	\$2.00
Hawthorn	\$3.80	West Coast	\$1.45
St Kilda	\$1.22	Sydney	\$2.50
Port Adelaide	\$2.30	Carlton	\$1.70
Adelaide	\$1.55	Melbourne	\$2.00
Essendon	\$1.18	North Melbourne	\$1.20
Fremantle	\$4.25	Brisbane	\$4.00

To see how this works, consider the first game. Suppose the bookmaker assigns a probability of p to a win by Collingwood (and thus a probability of $1-p$ to the event that Richmond wins).[3] Suppose I then come and wager \$1 on a Collingwood win. If the bookmaker has chosen p correctly, then I stand to win

[1] See "Heads I win, Tails you lose" by G A Watterson in *Function, Vol* 1, *Part* 3, p. 9.
[2] And, for simplicity, ignore the possibility of a draw.
[3] How this is done is, of course, a trade secret. The News-Item in *Function* was concerned with exactly this point: what is the best way to do this?

\$2.30 with probability p. The bookmaker will want my expected profit to be negative, and so chooses p to make $2.30p < 1$. Similarly $1.55(1-p) < 1$.

Now make the two expressions on the left-hand sides of these inequalities equal to one another; we will see why in a minute. We then have

$$2.30p \;=\; 1.55(1-p)$$

and this tells us that (to 2 decimal places) $p = 0.40$. We now have $2.30 \times 0.40 = 0.92$ and $1.55 \times 0.60 = 0.93$, where there is a small rounding effect in the second decimal place.

So we can deduce (subject to justifying the equality assumption) that the bookmakers involved with National Sportsbet assessed the probability of a Collingwood win at 0.40. For each dollar wagered on Collingwood they reckoned to outlay 92 cents, and for each dollar wagered on Richmond they thought to hand back 93 cents. Either way, they stood to make a profit of about 8%.

Now, actually, I've presented the calculations backwards. What the bookmaker does first is to decide on the value of p. The other given is the extent of the bookmaker's profit: around 8%. What is then calculated is the dividend paid in the event of a win by either team. So, if the bookmaker thought that the probability of a Collingwood win was 0.4, then the dividend to be paid out if Collingwood wins is \$(0.92/0.4), i.e. \$2.30. The dividend to be paid out in the event of a Richmond win is \$(0.92/0.6), i.e. \$1.533333…, an amount rounded to \$1.55.

The bookmaker is assured of a profit of about 8%, whichever team wins (assuming always that the initial estimate of p is correct). This is why, earlier on, we equated the two expressions; had one been significantly larger than the other, then the profit margin in the case of a Collingwood win would be significantly different from that applying in the event of a Richmond win. What the bookmaker aims to do is to ensure the 8% return, *whatever the outcome.*

So in the general case where D_A is the dividend paid out (per dollar invested) to a punter who successfully backs Team A, and D_B the dividend paid out (again per dollar invested) to a punter who successfully backs Team$B$, on each dollar bet the bookmaker pays $\$D_A$ with probability p and $\$D_B$ with probability $1-p$. Thus if Team A wins the expected payout is $\$pD_A$, while if Team B wins the expected payout is $\$(1-p)D_B$. The aim is to choose p in such a way as to *ensure* a profit margin m in each case. So we have

$$pD_A = (1-p)D_B = 1-m.$$

From the first of these equalities

$$p = \frac{D_B}{D_A + D_B} = \frac{\dfrac{1}{D_A}}{\dfrac{1}{D_A}+\dfrac{1}{D_B}}$$

(The apparently clumsy form of the second of these expressions is given here as this is the form that generalises to cases where more than two outcomes are possible. You may care to analyse this as an exercise.)

The bookmaker's expected profit also follows from the analysis given above. We find

$$1-m = \frac{D_A D_B}{D_A + D_B} = \frac{1}{\dfrac{1}{D_A}+\dfrac{1}{D_B}}$$

Because m is quite small (about 0.08 as we have seen) we may simplify this formula by using the approximation $1-m \approx \dfrac{1}{1+m}$ to reach the simplification

$$1+m \approx \frac{1}{D_A}+\frac{1}{D_B}$$

This gives us a very simple way to deduce the bookmaker's odds from the published dividends which can now be illustrated by the above case. Take the *reciprocals* of the dividends and add them.

$$\frac{1}{2.30}+\frac{1}{1.55} = 1.08.$$

If we now take each of the reciprocals and divide by this number, we obtain the bookmaker's estimate of p. In this case $\dfrac{1}{2.30} = 0.435$, and if this number is divided by 1.08, the result is $p = 0.4$. Notice how the 8% profit shows up in the course of this calculation. Readers may care to deduce the odds that National Sportsbet assigned to each of the games in AFL Round 4 and the profit margin (always near 8%) allowed for each game.

Now let us look at the situation described by Dr Watterson in the *Function* article referred to earlier. This began by considering an actual case where a set of dividends was offered that had been arrived at without due consideration. The details were topical then but need not detain us today. Suffice it to say that the dividend to be collected (per dollar invested) if Team A won was \$1.40, and if Team B won then the punter stood to collect \$4.50 on each dollar invested.

If we consider the sum of the reciprocals of these numbers, we find

$$\frac{1}{1.40} + \frac{1}{4.50} = 0.9365\ldots$$

and this number is *less than* 1. The bookmaker's profit from such odds is thus *negative*. Let us see how we could make money from this situation.

The idea is to invest \$$M$ on the first outcome and \$$N$ on the second. Thus if Team A wins we get back \$1.40M, and if Team B wins we get back \$4.50N. We want both of these receipts at least to return the \$(M+N) we ventured. So we want both

$$1.40M \geq M + N$$
$$4.50N \geq M + N$$

to hold. These inequalities simplify to[4]

$$2.50 \leq \frac{M}{N} \leq 3.50.$$

So if we wager \$$M$ on Team A and \$$N$ on Team B and make sure that the various amounts satisfy these inequalities, then we can't lose!

The previous article examined the case $M = 70, N = 22$. In this case the punter has outlayed \$92. If Team A wins, the return is $\$70 \times 1.40 = \$98$, for a profit of \$6; if Team B wins the return is $\$22 \times 4.50 = \$99$, for a profit of \$7. Thus the

[4] The work is left as an exercise for the reader.

punter must, on these figures, make at least \$6 for an outlay of \$92. The assured profit is 6.52%.

Now the question arises as to how we might maximise this assured return. Here again we use the equality principle: we set the assured return in the event that Team A wins equal to the assured return in the event that Team B wins. The reasoning is as before. Higher returns in the one case are counterbalanced by lower returns in the other.

Look first at one extreme case, $M = 2.50N$. Put $M = 5$, $N = 2$. Now if Team A wins, the punter scores $\$5 \times 1.4 = \$7$, a return of precisely zero on the initial outlay. If Team B wins, the return is $\$2 \times 4.50 = \$9$, and this is a profit of \$2 on an initial outlay of \$7, that is to say about 28.6%. At the other end of the scale, we may put $M = 3.50N$, and consider the case $M = 7, N = 2$. Now if Team A wins, the punter scores $\$7 \times 1.4 = \$9.80$, for a profit of \$0.80 on an initial outlay of \$9 (8.89% profit), but if, on the other hand, Team B wins the return is $\$2 \times 4.50 = \$9$, and the profit is zero. At either of these extremes, our *assured return* is actually zero.

But suppose we try to choose M and N so as to compromise between these two extremes. If the assured profit from one outcome is less than that from the other, then it is only the lower figure that we can count upon. Thus in this case, we set the two expected returns equal to one another. That is to say, we impose the condition that

$$1.4M = 4.5N.$$

So now we may set up an optimal strategy. Invest \$59, and put \$45 of this on Team A and \$14 on Team B. Either way, our return will be $\$1.4 \times 45 = \$4.5 \times 14 = \$63$. This is a return of \$4 on our outlay of \$59. We are assured of this \$4 profit, which represents a return of 6.78%.

We can trust to our luck by using other figures, and we *may* do better, but on the other hand we *could* lose out. The 6.78% is the best result we can be assured of achieving.

A few general remarks in conclusion.

Let us now look again at the general case where the dividends are D_A if we successfully back Team A or else a dividend D_B if we successfully back Team B. The condition that we can successfully take the bookmaker down is

$$\frac{1}{D_A}+\frac{1}{D_B}<1,$$

and our optimal strategy is to choose to invest $\$(M+N)$ with $D_A M = D_B N$. In this case, the return is $\left(\frac{1}{D_A}+\frac{1}{D_B}\right)^{-1}$ on every dollar outlayed, and by the inequality above this is clearly a profitable investment.

Similar remarks apply to (e.g.) horseraces, in which there are more than two "teams" competing. This is something I leave it to the reader to investigate (but see the brief remarks above). However, there are two points to consider. Horseraces still use, in many instances, "odds". These are to be used to calculate the dividends. Odds of 7:1 (against) for example give a dividend of \$8 for a \$1 stake; odds of 10:1 *on* give a dividend of \$1.10 for a \$1 stake, and so on.

Tote dividends are recorded *as dividends*, but the punter can't know until *after* the race what the actual dividend was, by which time the tote has already taken the profits out. The real world can be (indeed *is*) rather more complicated than the simplest models we construct. Nonetheless, until we understand the simple cases, we have no hope of following the more complex situations.

* * * * *

To make mathematics you must be interested in mathematics. The fascination of pattern and the logical classification of pattern must have taken hold of you. It need not be the only emotion in your mind; you may pursue other aims, respond to other duties; but if it is not there, you will contribute nothing to mathematics.

—W W Sawyer in *Prelude to Mathematics*

* * * * *

THE INTERACTION OF MATHEMATICS AND SCIENCE: AN EXAMPLE FROM OPTICS

Bert Bolton, University of Melbourne

Mathematics is often said to be useful, but there is something slightly condescending about the word 'useful'. Very often the mathematical analysis of a scientific problem reveals its essential nature, which is made clear when the problem is put into mathematical terms. As an example, one of the classic problems in astronomy starts with the measurements of the positions of the planets round the sun by Tycho Brahe (Danish, 1546–1601); such measurements, unrelated to any theoretical ideas, are called *empirical*. Brahe's measurements were then used by Johannes Kepler (German, 1571–1630) to show that the shapes of planetary orbits are ellipses. Finally Isaac Newton (English, 1642–1727) deduced mathematically the shape of these ellipses from the universal inverse square law of gravitational attraction between the sun and each planet. This and other mathematical discussions of scientific principles or fundamental laws make it tempting to say that mathematics is the right language in which science should be expressed, superseding the initial empirical observations. In fact science needs the empirical observations, experimental results based on theory and mathematical analysis of the underlying theory. There is a well-known example in the understanding of light that can be followed at both the empirical and the mathematical level. The mathematics involved is mostly straightforward geometry, but it illustrates the point of the theory and reveals a deeper understanding of nature, brought about by the mathematics.

The example concerns the reflection and refraction of light. These are phenomena which arise when light falls on surfaces such as glass. In Figure 1, a ray of light starts from a point A and proceeds to a point C via a reflection from a mirror. We may think of A as a light-source such as a lamp and C as a human eye. The eye sees A as reflected in the mirror at B. The mirror will be taken to be flat and the line BO is perpendicular to its surface. The angle ABO is called the *angle of incidence* and is represented by i; the angle OBC is called the *angle of reflection* and is represented by r. The lines AM and CN are drawn perpendicular to the mirror, and all the points are supposed to lie in one plane: that of the diagram. More will be said on this later; for the moment, we can say that this is what seems to happen empirically. There is no reason yet to say that B *must* lie in the plane defined by AM and CN, but there is no harm in taking this as a simple starting point.

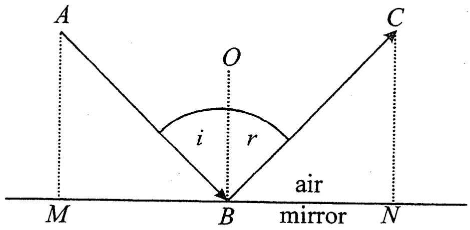


Figure 1

It has long been known empirically that the angle of incidence i is equal to the angle of reflection r. There have always been reflections in water surfaces and when mirrors were devised reflections in them could be studied. Some people have speculated that the early Egyptians used mirrors at the corners of their tunnels to reflect sunlight into the inner chambers of their pyramids. (Because of the rotation of the earth, the reflected light from the first, external, mirror would need to be rotated—almost certainly by hand—to direct the reflected sunlight onto the second mirror at the end of the first tunnel.)

Empirically, therefore, we take it that $i = r$ and for many situations that is all we need to know. But let us go further and look at Figure 1 mathematically. Consider the case in which B is not necessarily chosen to make $i = r$. But then if the point of reflection B were such as to make $i \neq r$, the path the light would travel would be longer than that taken if $i = r$.

To see this, draw Figure 2, where CN has been extended to C' with $CN = NC'$, and OB has been extended to O'. Notice that

$$\angle O'BC' = \angle OBC = r.$$

Notice also that the distance travelled by the light, ABC, is equal to the distance ABC'.

Very clearly, this distance will be minimised when A, B, C' lie in a straight line. But the condition for this is precisely that $i = r$.

Now we may also note that if B had been chosen to lie *outside* the plane defined by AM and CN the path from A to B to C would also have been lengthened. By drawing Figure 1 as we did, we chose the minimum length for the path of the light.

Light travels with a very large, but nonetheless finite, velocity, so another way to express the minimum principle is to say that the light ray gets from A to B to C in the shortest possible time.

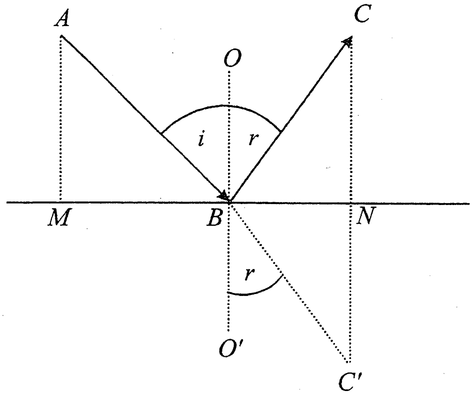


Figure 2

We thus have two possible interpretations of the reflection data. One is that light so travels as to minimise the total distance covered; the other is that light so travels as to minimise the total time taken. Fortunately we have further empirical data to distinguish between these two possible interpretations; it turns out that it is the time which is minimised.

Figure 3 shows the example of light being refracted from a source A, which we may think of as being situated in air, and reaching C, which we may think of as a fish eye in water. The word *refraction* (describing the bent path of the light ray) comes from a Latin word meaning "breaking". It relates to our observation of (say) a long pencil dipped into a glass of water. If we look down on it we see an apparent "break" in the line of the pencil. I have chosen the same symbols for the points in Figure 3 as I used in Figure 1. The angle i is still referred to as the angle of incidence, but r now is called the *angle of refraction*. The empirical law of refraction was given by Willebrord Snell (Netherlands, 1580-1626), and it states that

$$\frac{\sin i}{\sin r} = \text{constant}.$$

The constant in this equation is called the *refractive index* and it is often represented by the Greek letter μ.

For light leaving air and entering water, the value of μ is about 1.33. It is now known that this μ is the ratio of the speed of light in air, c_a, to the speed of light in water, c_w. Thus Snell's law for a ray of light leaving air and entering water is

$$\frac{\sin i}{\sin r} = \mu = \frac{c_a}{c_w}$$

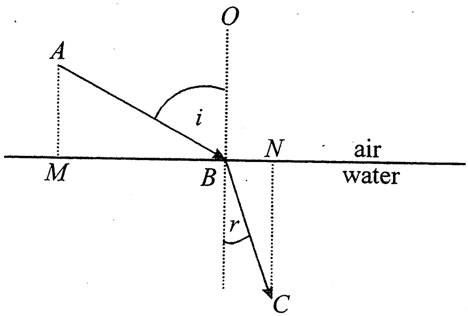


Figure 3

This form of Snell's law is equivalent to the light taking the shortest *time* in travelling from A to C. (Clearly *distance* is not minimised, as the line ABC is not straight!)

To see this, put $MB = x$, $MA = a$, $MN = l$ and $NC = b$. Our aim is to choose B (i.e. the value of x) in such a way as to minimise T, the total time taken by the light in travelling from A to C. But now

$$T = \frac{\sqrt{a^2 + x^2}}{c_a} + \frac{\sqrt{b^2 + (l - x)^2}}{c_w}$$

In order to minimise T, we need to set its derivative equal to zero. This gives

$$\frac{dT}{dx} = \frac{x}{c_a\sqrt{a^2 + x^2}} - \frac{l - x}{c_w\sqrt{b^2 + (l - x)^2}} = 0.$$

This equation may be rewritten as

$$\frac{x}{c_a\sqrt{a^2+x^2}} = \frac{l-x}{c_w\sqrt{b^2+(l-x)^2}}$$

and the left-hand side of this equation is the sine of the angle *MAB*, divided by c_a. Similarly the right-hand side is the sine of the angle *NCB*, divided by c_w.

But this is exactly the same as saying that

$$\frac{\sin i}{c_a} = \frac{\sin r}{c_w}$$

which is Snell's law.

This discussion has shown how valuable mathematics is for science. If we wish, we can keep our understanding of the natural world at the level of empirical observation and generalisations such as the laws of reflection and Snell's law of refraction. But showing that empirical data can be deduced mathematically introduces us to the possibilities of new experiments based on new ideas. For example, what of the values of c_a and c_w? Do they vary with density or temperature? Do they depend on the structure of the medium? These and other questions are not just empirical but are based on interpretations of the nature of light and its interaction with matter. Mathematicians need never think that what they discuss will be too 'pure' or abstract to be useful in science. The ideas of pure mathematics have constantly been found to be valuable in new fields of science. Many areas of physics have moved beyond just handling experimental data and are using powerful mathematical ideas. New problems are continually appearing; there is room for many workers.

A problem completely analogous to Snell's law of refraction may be put in the context of a beach drama. Let *MN* in Figure 3 represent the water's edge, and suppose that you are a surf lifesaver at *A*. You see a swimmer in difficulties at *C* in the water. You can run faster on sand than you can swim in water. The point is to minimise the time you take to get from *A* to *C*. That path will be given by Snell's law.[1]

* * * * *

[1] This problem is similar to one recently discussed in *Function* (see *Vol* 22, *Part* 5, pp. 147–153). However, it is not the same problem, as the reader may verify. The earlier problem has a closer optical analogue in the problem of "total internal reflection", but the details are omitted here.

A SUM RESULTING IN π

Mark Nolan, Year 9, St Leonard's College

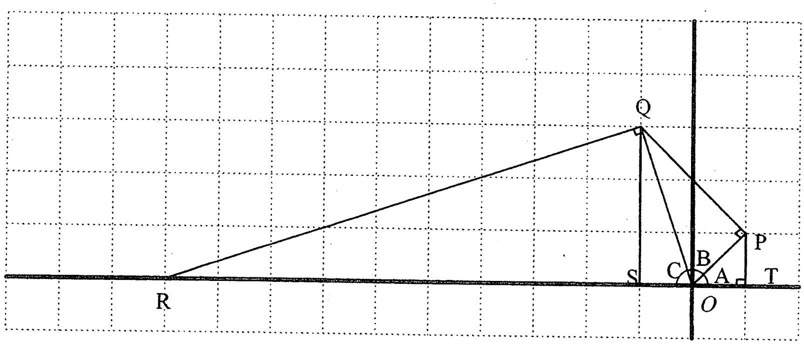
In this problem, I have to prove that

$$\mathrm{Tan}^{-1}1 + \mathrm{Tan}^{-1}2 + \mathrm{Tan}^{-1}3 = \pi .$$

[By $\mathrm{Tan}^{-1}x$ is meant "the (acute) angle whose tangent is *x*". It is also understood that all angles are to be measured in *radians*.]

As you can see in the diagram, there are three angles: Angle A, Angle B and Angle C. Angle A is a forty-five degree angle (although in this problem it is measured in radians and will be known as $\frac{\pi}{4}$), and so its tangent is 1. Angle B has been made to have a tangent of 2, and Angle C a tangent of 3. Together, these add to π.

To prove this, we use the diagram below. There are three special points on this diagram; P, Q and R. P is the point such that the angle between OP and the *x*-axis has tangent 1; angle POQ has tangent 2; angle QOR has tangent 3. I have to prove that R is lying exactly on the *x*-axis, which will show that the three angles add to π. To do this, I will first have to work out the coordinates of the other points. It is possible to figure this out by using only logic and Pythagoras' Theorem.



It is quite simple to find the coordinates of P. It is at (1,1). This is because, as mentioned, A is an angle of $\frac{\pi}{4}$.

The next point to find is Q. The line PQ has been made to be twice as long as the line OP. Using Pythagoras' theorem, I found that the line OP has a length of $\sqrt{2}$ and therefore the line PQ has a length of $2\sqrt{2}$, or $\sqrt{8}$. Using Pythagoras' theorem once again, I find that the line QO has a length of $\sqrt{10}$. I already know the coordinates of the points O and P, and their distances from Q, so I should be able to find the coordinates of Q.

Let us presume that the coordinates of Q are (x, y). To find out exactly what these pro-numerals stand for, I must discover the relationship between x and y. I found a very useful equation simply by using Pythagoras.

I saw that $\sqrt{x^2 + y^2} = \sqrt{10}$ because I already knew that the line QO had a length of $\sqrt{10}$. I then simplified the equation even more to get $x^2 + y^2 = 10$. This did not fully solve everything, but it was a vital clue. Using the same method, I saw that

$$\sqrt{(x-1)^2 + (y-1)^2} = \sqrt{8},$$

which I later simplified to $(x-1)^2 + (y-1)^2 = 8$. I then expanded this equation, giving me

$$x^2 - 2x + 1 + y^2 - 2y + 1 = 8.$$

After rearranging this statement into a more orderly fashion, I came up with

$$x^2 + y^2 - 2(x + y) = 6.$$

Now I could use the equation I discovered earlier to show that $10 - 2(x + y) = 6$. This obviously shows that $x + y = 2$.

Now that I know this, I can see that $y = 2 - x$. I now have found a very simple relationship between x and y. If I replace every y I come across with a $2 - x$, I can solve the mystery of the coordinates to the point Q! Instead of saying $x^2 - y^2 = 10$, I could say that $x^2 + (2-x)^2 = 10$. This equation has the solutions $x = 3$ and $x = -1$. From the diagram, we see that it is the second of these solutions we want. When $x = -1$, $y = 3$.

I have now found the coordinates of the point Q to be $(-1,3)$. With this knowledge, and using the same method as I have just used, I will attempt to find the coordinates of the point R. If it lies directly on the x–axis, then it will prove that the angles A, B and C do add up to π. If the point R is slightly above or below the x–axis, then I am wrong.

This is how I am going to find the coordinates of the point R. It is quite similar to the way in which I found the coordinates of the point Q. So far, I know that $Q = (-1, 3)$ and that $RQ = 3\sqrt{10}$. I will call the point R (x, y).

I will make a statement using Pythagoras that is similar to the one I first used when I was figuring out the coordinates of Q. It says that $(x+1)^2 + (y-3)^2 = (3\sqrt{10})^2 = 90$, which means that the line RQ has a length of $3\sqrt{10}$, or $\sqrt{90}$. We already know that the line QO has a length of $\sqrt{10}$, so by using Pythagoras yet again, we shall see that the length of the line RO is $\sqrt{100}$, or 10. This does not quite prove that $y = 0$, but it narrows down the search a bit because it tells us that $x^2 + y^2 = 100$.

So now we have discovered a relationship between x and y, which is that $y^2 = 100 - x^2$. We can replace every y^2 with this expression, and it will make this problem a lot easier. We can expand an earlier discovery:

$$(x+1)^2 + (y-3)^2 = 90 \quad \text{to give} \quad x^2 + 2x + 1 + 100 - x^2 - 6y - 9 = 90.$$

This can now be simplified to: $x = 3y - 10$.

Now we have another relationship between x and y! Since we already know that $x^2 + y^2 = 100$, it is obvious that $(3y-10)^2 + y^2 = 100$. This shows that y could be either 0 or 6. This is because the statement above was found due to the fact that it showed the point which is a distance of 10 units from the point O and also $\sqrt{90}$ units from the point Q. But there are two points which satisfy this criterion. One is at $(-10, 0)$. Now I have shown that

$$\text{Tan}^{-1}1 + \text{Tan}^{-1}2 + \text{Tan}^{-1}3 = \pi\ !$$

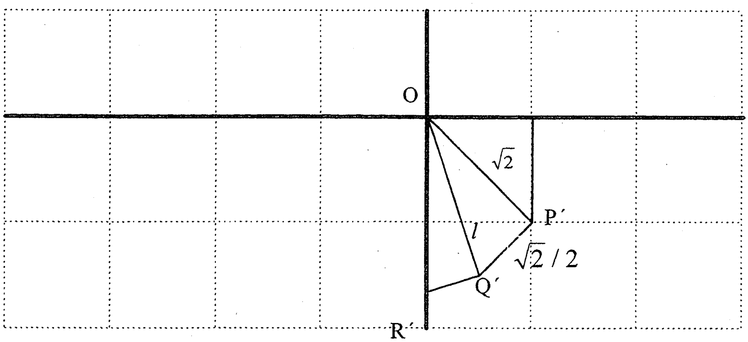
But it seems that I didn't really need to go to all that trouble. There was actually an easier way to have proved this statement.

So far, I haven't yet mentioned the points S and T which you can see in the diagram on the first page of this article. S is at the coordinates $(-1,0)$ and T is at the coordinates $(1,0)$.

As soon as I had found the location of the point Q, I could have used this procedure to solve the problem without having to find the coordinates of the point R. If I had drawn a vertical line from S to Q, just after I had discovered that Q was located at $(-1,3)$, I would have seen that the distance between S and O was 1 unit and the distance between S and Q was 3 units. This would immediately show that the angle $\angle SOQ$ was equal to $\mathrm{Tan}^{-1}3$. Since the point S has already been made to be on the x–axis, the problem is already solved!

I conclude from my findings that $\mathrm{Tan}^{-1}1+\mathrm{Tan}^{-1}2+\mathrm{Tan}^{-1}3=\pi$.

This next diagram shows another way to reach this formula.



1/3

In this problem, I have to prove that $\mathrm{Tan}^{-1}1+\mathrm{Tan}^{-1}\frac{1}{2}+\mathrm{Tan}^{-1}\frac{1}{3}=\frac{\pi}{2}$. To show this, I will use coordinate geometry. Using radians, the three angles shown in the diagram add up to a right angle, or $\frac{\pi}{2}$. If I can show that the point R' is on the y–axis, I will have proved this.

Obviously, the point P' is at $(1,-1)$. Because the point Q' is half the distance from P' as P' is from O and the angle $OP'Q'$ is a right angle, Q' is at $(\frac{1}{2},-\frac{3}{2})$.

Call the length of OQ′ equal to l. If I use Pythagoras' theorem, I can find the length l is $\sqrt{\frac{5}{2}}$, and so the length of the line R′Q′ is $\frac{1}{3}\sqrt{\frac{5}{2}}$. If I were to call the point R′ (x, y), then because of this fact, we have

$$\left(x-\frac{1}{2}\right)^2+\left(y+\frac{3}{2}\right)^2=\frac{5}{18}.$$

If I use Pythagoras' theorem again, I can find the length OR′ is $\sqrt{l^2+\left(\frac{l}{3}\right)^2}$ and this works out to be $\frac{5}{3}$. So now I have another equation

$$x^2+y^2=\frac{25}{9}.$$

These two equations can be solved in the same way I solved pairs of equations before. We can find out that $x=3y+5$ and so $(3y+5)^2+y^2=\frac{25}{9}$. When I solved this equation, I found that $y=-\frac{4}{3}$ or else $y=-\frac{5}{3}$. Which is correct?

To find out, I looked at the coordinates of Q′. These are $(\frac{1}{2},-\frac{3}{2})$, so that Q′ is $\frac{3}{2}$ units down the page from the x-axis. R′ has to be further down still and so its y-coordinate must be $-\frac{5}{3}$, not $-\frac{4}{3}$. So now I know that $y=-\frac{5}{3}$ is the value I must use. But this gives $x=0$, which means that the point R′ is on the y–axis, which is what I wanted to prove!

I have proved that $\mathrm{Tan}^{-1}1+\mathrm{Tan}^{-1}\frac{1}{2}+\mathrm{Tan}^{-1}\frac{1}{3}=\frac{\pi}{2}$.

This is actually related to the previous formula

$$\mathrm{Tan}^{-1}1+\mathrm{Tan}^{-1}2+\mathrm{Tan}^{-1}3=\pi$$

because $\mathrm{Tan}^{-1}\frac{1}{2}=\frac{\pi}{2}-\mathrm{Tan}^{-1}2$ and $\mathrm{Tan}^{-1}\frac{1}{3}=\frac{\pi}{2}-\mathrm{Tan}^{-1}3$. So the two formulas are really the same.

So I have now proved the formula $\mathrm{Tan}^{-1}1+\mathrm{Tan}^{-1}2+\mathrm{Tan}^{-1}3=\pi$ in three different ways.

[Editors' Note: The formula $Tan^{-1}1+Tan^{-1}2+Tan^{-1}3=\pi$ *was pointed out to us by Dr Mike Englefield, who found it in a recent issue of the American Mathematical Monthly. Neither he nor we had seen it explicitly given in this neat form. We suggested to Mark that he look at the formula and try to prove it. The article above is the result of Mark's efforts. The formula*

$$Tan^{-1}1+Tan^{-1}\tfrac{1}{2}+Tan^{-1}\tfrac{1}{3}=\tfrac{\pi}{2}$$

is equivalent to the first, as Mark points out. This latter formula may be somewhat simplified. Because $Tan^{-1}1=\frac{\pi}{4}$, *it may be written as*

$$Tan^{-1}\tfrac{1}{2}+Tan^{-1}\tfrac{1}{3}=\tfrac{\pi}{4},$$

and in this form it is reasonably well-known. Indeed it has appeared in Function *before. In an article entitled "Pi through the Ages", J M Howie quoted it and referred to it as "beloved by sixth form examiners the world over" (see* Vol *4,* Part *1,* p. *9).*

Other proofs are possible. For instance, there is a general rule

$$Tan^{-1}x+Tan^{-1}y=Tan^{-1}\left(\frac{x+y}{1-xy}\right)$$

and if we insert the values $x=\frac{1}{2}$ *and* $y=\frac{1}{3}$ *into this result, we find*

$$Tan^{-1}\tfrac{1}{2}+Tan^{-1}\tfrac{1}{3}=Tan^{-1}1=\tfrac{\pi}{4}.$$

This same general rule could also be used in other ways. Mark's three proofs however give considerable geometric insight into the formula.]

* * * * *

NEWS: Rare Maths Book sold at Auction

The Greek mathematician Archimedes lived from about 287BC to 212BC, when he was killed during the siege of Syracuse. He is regarded as one of the very greatest mathematicians who ever lived, and his contributions to geometry, to mechanics and to what (much later) became calculus have ensured this honoured place for him in the history books.

Many of his writings survive and are available to modern readers. But it is important to understand just how this has happened. Back when Archimedes lived, and for over a thousand years after that, books all had to be copied by hand. Copies were made of copies of copies ..., and some of these various copies were better than others. Others came to sad ends as a result of fire, wars, vandalism and their like. Eventually the surviving copies found their way into the rare book rooms of major libraries, such as the different national or university collections. Modern editors try to compare all the surviving copies (a very small subset of the total!) and so to reconstruct the original as best they can.

It is therefore a great day when a new copy comes to light, and this is what happened recently with the texts of two of Archimedes' best works: *On Floating Bodies* and *The Method of Mechanical Theorems*. The first of these is concerned with Hydrostatics, the science of fluid equilibrium, and it is the one that contains (as its Proposition 7) the famous "Archimedes' Principle", which in a modern version says that if a solid body floats in a liquid, then the amount of liquid it displaces weighs precisely the same as the solid does. The second is in some ways more interesting work in that it uses the methods of Statics (the general science of equilibrium) to come up with geometrical results. For example its Proposition 2 determines the volume of the sphere. Later on, Archimedes got these same results by more efficient, straightforwardly geometric, means. But his earlier proofs tell us a lot about how the mind of this great genius worked.

It is believed that the parchment onto which the works were copied dates from the 12th Century AD, and so it is very far removed in time from Archimedes himself. It has survived largely unnoticed because it is a *palimpsest*. [This technical word refers to a manuscript written over the top of another. In this case, there was an attempt to erase the original Archimedean text and a later text (here devotional material from an Eastern branch of the Christian Church) superimposed on the erasure.] However it is often possible still to read the original text, especially with 20th Century techniques like the use of ultraviolet light. In this case, preliminary examination shows that quite a lot can be read, and that the texts are almost complete, though some pages and words near the margins are missing.

The work came to light only recently, when it was offered for sale at Christie's (the well-known auction house) last October. Christie's were selling it on behalf of its owners whom they described as an anonymous French family. They satisfied themselves that this family had held the manuscript for "three quarters of a century" and that they had acquired it legitimately by buying it (in or about 1920) from the previous owners: the Greek Orthodox Patriarchate of Jerusalem, who seem themselves to have acquired it (well before this) from a library in Constantinople, now Istanbul.

There was some considerable discussion before the auction took place over the rightful ownership of the document, but a legal action brought by the Patriarchate in the Federal Court of New York failed, and so the auction was allowed to proceed. It was suggested that Princeton University in the USA might bid for it, but they did not do so, yielding to the desire of the Greek government to acquire the work.

When it came to the auction, bidding began at $US480,000 and in a matter of minutes ran up to $US2,000,000, at which price it was knocked down. The successful bidder was Simon Finch, a London dealer, who often acts as an agent in such matters. The actual buyer has not been identified, but is described as an anonymous American collector. A rumour that it was Bill Gates has however been officially denied.

The Greek government therefore missed out, but the Patriarchate have not ruled out taking their legal claim further. This would entail close examination of the way in which the French family acquired the work. Did the person who sold it to them actually have the authority to do so? That sort of question.

Meanwhile the good news is that the new owner has declared that the work will be made available to reputable scholars, so that its significance for the understanding of Archimedes and his work can be fully assessed. This will be a painstaking and lengthy process, but in due course we may hope to have a better text for these two works and an increased understanding of Archimedes' thought.

* * * * *

HISTORY OF MATHEMATICS

The Isoperimetric Problem

Michael A B Deakin

If we have a length of string, say, with its two ends joined together, we can so place it on a flat surface, like the top of a table, that it encloses an area A, say. Suppose that the length of the string is L. How should we arrange the string so that without altering its length, we ensure that the area A is maximised? This problem is the so-called "isoperimetric problem", or "classic isoperimetric problem", and it is one of a class of similar problems which collectively are also known as "isoperimetric problems". The word "isoperimetric" comes from *isos-*, meaning "equal" and *perimeter*, which refers to the length of the boundary of the enclosed area A.

The history of such problems goes back a very long way; we could even say to a time before reliable history. Legend has it that the famous queen Dido was shipwrecked with her companions on the coast of North Africa and asked the local residents for a grant of land where they could settle. They allowed her as much land as could be "encompassed by the hide of an ox". Dido cut the oxhide into very thin strips and joined these together to form a long strip of leather. We shall see in detail later on quite how she used this strip to enclose quite a large region of land, upon which she built the city of Carthage. In time this became a large and powerful centre of trade and agriculture, which in its heyday challenged the might of Rome. (Rome however prevailed and Carthage was reduced to rubble; its ruins lie in the modern state of Tunisia.)

More certainly, the Greek mathematician Zenodorus produced a book *On Isometric Figures*, probably some time between 200 BC and 90 BC. Little is known of Zenodorus, and his book no longer survives in the form in which he wrote it. However portions of it are still with us because later mathematicians reproduced extensive passages from it in their own works. For example, the much later Pappus (4th century AD) proves a number of theorems on isoperimetric problems in Book V of his *Collection*, and his source is Zenodorus. Other parts are preserved in the writings of Pappus's somewhat younger contemporary, Theon of Alexandria.[1] Because we no longer have the original work, it is not always clear, however, quite what theorems that are still available to us are due to Zenodorus and which to Pappus.

[1] Theon was the father of Hypatia, whose story was told in *Function, Vol* 16, *Part* 1.

However, those theorems are:

(1) Of all regular polygons of equal perimeter, that with the largest number of angles encloses the greatest area.

(2) A circle encloses a greater area than any polygon with the same perimeter (circumference).

(3) Of all polygons with the same number of sides and the same perimeter, the regular polygon encloses the greatest area.

Pappus added a fourth theorem, not proved by Zenodorus:

(4) Of all segments of a circle having the same perimeter, the semicircle has the greatest enclosed area.

Let us have a look at the proof of the second of these theorems. Archimedes had proved that the area of a circle is half the product of its circumference and its radius. That is to say that if r is the radius, then

$$A = \frac{1}{2}(2\pi r)\,r$$

The proof uses this fact and takes a *regular* polygon and a circle with the same circumference (perimeter) as this polygon. But now if we compare the radius of *this* circle with the distance from the centre to the midpoint of a side, we find that the latter is the lesser, so that the area of the polygon is correspondingly less. Readers may care to recast this proof into modern form using trigonometry. In order to complete the proof we need the third of the three theorems given above in order to justify the use of regular polygons in the calculation.

The matter that perhaps most caused interest in Pappus's account was his deduction from these theorems that of all the shapes that pack together, the regular hexagon encloses the greatest area for the length of its perimeter. Pappus used this to demonstrate the wisdom of bees, whose honeycombs adopt the hexagonal shape. (Bees were very important economically in his time as they provided the principal, almost the only, source of sweetening.)

Now let us get back to Dido's problem of securing for herself and her companions the greatest area of land that she could. Here is how modern authors tend to interpret the ancient legend.

We have a straight piece of seacoast and we are so to position the strip of leather as to join two points on this and to be arranged inland in such a way as to maximise the enclosed area A. See Figure 1.

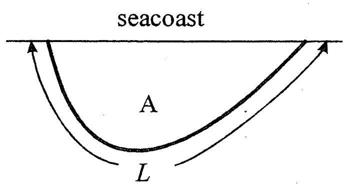


Figure 1

I now show that the curve adopted by the leather strip should be semicircular. (This is similar to the fourth of the theorems listed above, but is not the same, because the length of the seacoast is here not included in the total; only the length of the leather strip.)

To show that the required maximum is given by a semicircle, take a curve which is not a semicircle as shown in Figure 2.

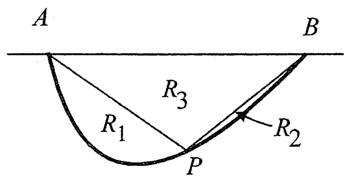


Figure 2

A semicircle has the property that if *any* point P on it is joined to the ends A and B as shown, then the angle APB is a right angle. Moreover the semicircle is the *only* curve joining A and B which possesses this property. Because the curve in Figure 2 is *not* a semicircle, there must then be some point P such that angle APB is *not* a right angle. Figure 2 has chosen such a point as the location of P. Join AP and BP. We now have three regions R_1, R_2, R_3, as shown. Now, *without altering R_1 or R_2 in any way,* adjust the area R_3 by making angle APB equal to a right angle. This necessarily increases R_3 (prove this as a simple exercise) and so *increases* the total area. We may always so increase the area in this way unless *for every point P* the angle APB is a right angle. Thus if the curve is *not* semicircular, we may always make its area larger, and so the area is not maximised.

This elegant proof is much more recent than Queen Dido, Zenodorus, or even Pappus or Theon. It is due to the nineteenth century geometer Jakob Steiner, who published it in 1836.[2]

We are now in a position to solve the classical isoperimetric problem with which this article opened. The solution is: *Of all the simple closed*[3] *curves that may be constructed in a plane and having a total perimeter L the circle is the one that maximises the enclosed area A.*

Steiner likewise produced a nice proof of this, and I give it here. Before I do so, however, I remark that the theorems given by Zenodorus and Pappus go a long way toward proving the result, but not *all* the way (although we may use their work to build a proof that does do the full job).

Steiner's proof takes a simple closed curve C and chooses any point P on it. We now proceed a distance $L/2$ along the curve, and reach a point we will call Q. The line PQ now divides the interior of C into two regions R_1 and R_2. (See Figure 3.)

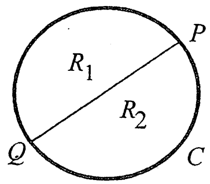


Figure 3

If the curve C is *not* a circle, then either one or other of these regions has the larger area, or else (what is less likely) they may just happen to have exactly the same area.

In the first case, we may increase the area with no corresponding increase in perimeter simply by replacing the smaller area with a mirror image of the larger, so

[2] Steiner has featured in earlier articles in *Function*. The idea behind this proof is very like that behind a proof given for one of the theorems in my history article in *Vol* 20, *Part* 3, which concerned yet another isoperimetric problem.

[3] A closed curve is one whose ends join up. A *simple* closed curve is one that doesn't intersect itself and so unambiguously encloses an area..

the total area cannot previously have been maximal. In the second case, for C not to be a circle, at least one of its halves will fail to be a semicircle. But if we replace the arc that is *not* a semicircle by one that *is*, we increase the area bounded by the arc and the line PQ (by the solution to Dido's problem). Now reflect this new semicircular arc in PQ and so reach a circle, which will be the solution to the problem.

These two proofs are so elegant and ingenious that it is a pity to have to point out that they are incomplete in one technical detail. This was not realised till a later mathematician, P G L Dirichlet (1805-1859) noticed it. His objection is really very subtle, but it is telling. To demonstrate its force, I will concentrate on Steiner's solution of the Dido problem. Because he used this in the later theorem on the classical isoperimetric problem, a flaw in this proof affects the two of them.

When Steiner discussed the Dido problem, what he showed was that if the arc AB was not semicircular, then it was not the curve that enclosed the greatest area. It could always be improved on. But to reach the semicircle by means of such improvements involved infinitely many steps (there are infinitely many possible points P in Figure 2), and so this is not a viable programme for getting to the optimal curve. In fact, Steiner didn't envisage doing this. Rather his proof assumes *that there is a curve that maximises the area* and then shows that no curve other than the semicircle could possibly be this maximising one.

As Dirichlet rightly pointed out, the assumption that the maximising curve actually exists is something that itself requires proof. The point can be met and the gap closed, although the details get too complicated to go into here, but perhaps I can indicate the need for such caution by referring to another problem. This is in the same family as those we have been looking at, but is a more distant cousin, if I may so put it.

This problem is known as the "Kakeya Problem" and it asks for the curve enclosing the minimum area inside which it is possible to rotate a straight line segment of length 1 through 360°. An equilateral triangle of side $2/\sqrt{3}$ can easily be shown to work. But could we not perhaps do better?

Well, yes we can, but only if we stop using "convex curves". Convex curves are those that always "bulge out" rather than inwards. If we wrap a thread around the shape made by a convex curve, the thread and the shape keep in contact at all points. Clearly in the case of the isoperimetric problems discussed up to this point, we want convex curves, as the enclosed area is to be *maximised*. But in the case of the Kakeya problem, the area is to be *minimised* so matters are different. It turns out that if we only look at *convex curves*, then the equilateral triangle is the best we can do with the Kakeya problem. The real surprise comes however when we allow non-convex curves.

In this case, there is *no answer*! Whatever curve we find, there is always another curve that does better. The area can be made to be arbitrarily small, arbitrarily close to zero. But we can't have zero itself as the enclosed area, because in this case we stop having a properly defined curve, and also the geometric shape that the successively better curves do converge upon in fact *doesn't* allow the required rotation.

So Dirichlet was right to point out the gap in Steiner's proofs. A pity really. It is also a pity that the filling in of this gap is a matter of some difficulty. It is certainly outside the scope of *Function*.

* * * * *

> At the age of 21 I experienced a second wonder ... in a little book dealing with Euclidean plane geometry, which came into my hands at the beginning of the school year. Here were assertions, as for example the intersection of the three altitudes of a triangle in one point, which—though by no means evident—could nevertheless be proved with such certainty that any doubt appeared to be out of the question. This lucidity and certainty made an indescribable impression upon me. ...
>
> If thus it appeared that it was possible to get certain knowledge of the objects of experience by means of pure thinking, this "wonder" rested upon an error. Nevertheless, for anyone who experiences it for the first time, it is marvellous enough that man is capable at all to reach such a degree of certainty and purity in pure thinking as the Greeks showed us for the first time to be possible in geometry.

—Albert Einstein

* * * * *

COMPUTERS AND COMPUTING

Parallel Computing

Peter Grossman

Most of the computers in use today, including the personal computers you are probably familiar with, are *serial* machines: they perform computations one operation at a time. In recent years, however, computers have been built which are capable of performing many operations at the same time. Such machines, comprising many processors which can operate simultaneously, are called *parallel* computers. A typical parallel computer might have $2^{16} = 65536$ processors, connected to each other in an arrangement corresponding to a 16-dimensional hypercube, in much the same way that $2^3 = 8$ points can be placed at the vertices of a (3-dimensional) cube. Each processor does its own computations, and sends and receives information to and from its neighbouring processors. Clearly, performing processes in parallel rather than serially has the potential to speed up computations enormously. Parallel computers are now being used in many areas, including vision systems for robots, searching large databases, simulating complex physical systems such as the atmosphere, and artificial intelligence.

However, there is more to parallel computing than just building a parallel machine. We cannot expect that a program written for a serial computer will automatically run faster on a parallel computer. If we want to be able to take advantage of a machine's ability to carry out several processes at once, we need to design our programs accordingly.

Here is a simple example. Suppose we want a program that inputs a list of numbers, and outputs the largest number in the list. (For simplicity, we'll assume the numbers are distinct.) There is a straightforward serial algorithm for doing this. It uses a variable, *max*, which is initially equal to the first number in the list. The algorithm begins by comparing *max* with the second number in the list, and if that number is larger than *max*, the value of *max* is updated to the larger value. Proceeding in this way through the list, each number in turn is compared with the current value of *max*, and the value of *max* is updated if necessary.

Here is the algorithm written in pseudocode:

Algorithm 1:

1. Input $a_1, a_2, \ldots, a_n$
2. $max \leftarrow a_1$
3. **For** $i = 2$ **to** n
   3.1. **If** $a_i > max$
      3.1.1. $max \leftarrow a_i$
4. Output max

Algorithm 1 is not the only way of finding the largest number in a list. For example, given a list of four numbers, we could find the greater of the first and second, then the greater of the third and fourth, and finally the greater of the two numbers obtained from the first two steps. This process is rather like a tennis tournament: the first two comparisons are the "semifinals", and the "winners" meet each other in the "final" (the last comparison).

The tournament method described above can be generalised to handle lists of any length. In the first round, the first and second numbers in the list are compared, then the third and fourth numbers are compared, then the fifth and sixth, and so on. The winner of each comparison proceeds to the second round. (If the list contains an odd number of numbers, the last number in the list proceeds to the second round without taking part in a comparison.) In the second round, the process is repeated using the list of winners from the first round, and so on until only one number remains.

Is there any advantage in using the tournament method instead of Algorithm 1? If the list contains four numbers, for example, both the tournament method and Algorithm 1 make three comparisons, so there is no advantage in this case. In fact, we can show that *any* comparison-based algorithm for finding the largest number in a list of n numbers must perform at least $n-1$ comparisons. The proof is not difficult; it runs as follows.

Suppose we have found the largest number in a list of n numbers. Then we know that each of the other $n-1$ numbers is *not* the largest number. Now, in order for us to know that a number is not the largest, it must have "lost" at least one comparison, i.e., it must have been compared with another number and been found to be smaller. (If a number had won all the comparisons it was involved in, or if it had not been involved in any comparisons, how could we be sure it wasn't the largest number in the list?) Since $n-1$ numbers have each lost at least one comparison, at least $n-1$ comparisons must have been performed.

Algorithm 1 makes $n-1$ comparisons, so we can't do any better with a different algorithm. On a serial machine, finding the largest number in a list takes at least $n-1$ times the time it takes to perform a comparison, in addition to the time taken by the other operations in the algorithm.

Now suppose we have a parallel computer at our disposal. Could Algorithm 1 take advantage of it by doing some of the comparisons simultaneously? A moment's thought reveals that the answer is no. Algorithm 1 needs to know the result of each comparison before it can proceed to the next one.

The tournament method is a different matter. Just as the matches in a round of a tennis tournament are often played simultaneously, so the comparisons in one round of the tournament method can be carried out in parallel. How much time would this algorithm take on a parallel machine? Since each round now only takes the time needed for one comparison, the answer equals the time for one comparison multiplied by the number of rounds (plus the time taken for the other operations, including passing information between processors, which we will ignore here). The number of rounds for a list of n numbers is about $\log_2 n$. (To be precise, it is exactly $\log_2 n$ if n is a power of 2, and it is the smallest integer greater than $\log_2 n$ otherwise.)

For example, finding the largest number in a list of 1000 numbers using a serial algorithm takes (at least) 999 times the time taken for one comparison, whereas solving the same problem on a parallel machine using the tournament method takes only 10 times the time taken for one comparison—about one percent of the time taken by a serial algorithm. For longer lists of numbers, the improvement is even more dramatic.

We have introduced parallel algorithms by looking at one particular problem: finding the largest number in a list. You might like to explore how algorithms for some other problems would work on a parallel machine. For example, what algorithm would you normally use to find the sum of the numbers in a list? Could it be made to run faster on a parallel computer? If not, could the approach taken in the tournament method be used to sum a list of numbers on a parallel computer? There are many other problems you could think about, such as searching for a number in a list, or (a more challenging problem) sorting a list of numbers into order. How could these be done as parallel computations?

* * * * *

PROBLEM CORNER

PROBLEM 22.4.1

Use a "balls in cells" approach to find the probability that there are no runs longer than two consecutive numbers when drawing 6 numbers from $1, 2, \dots, n$.

SOLUTION (Malcolm Clark)

Use the same balls-in-cells and line-of-symbols idea as described in *Function Vol* 22, *Part* 4, pp. 122–125, but this time define new symbols S (for single) and D (for double) as follows:

> S = single occupied cell followed immediately by an empty cell.
> D = two consecutive occupied cells followed immediately by an empty cell.

For the last S or D,

> Last S = single occupied cell.
> Last D = two consecutive occupied cells.

As in the article referred to above, a circle denotes a "free" empty cell, i.e., one not tied to an S or D. There are four ways in which there can be no runs longer than two in a selection of six numbers from 1 to n.

(a) *Six single numbers*. As shown in the article, the number of arrangements is

$$\frac{(n-5)!}{6!(n-11)!}$$

(b) *Four singles and one double.*

With the above definition of symbols, each possible arrangement of four singles and one double can be represented uniquely by an arrangement of $(n-5)$ symbols, of which there are 4 S's, 1 D, and $(n-10)$ circles. Out of the original n circles, we lose 4 which are linked to the first 4 S's or D's, plus one more for each D, in this case a total loss of 5 circles.

Hence the number of ways of getting four singles and a double is equal to the number of arrangements of $(n-5)$ objects, of which 4 are of one type (S), 1 is of another type (D), and the remainder are of a third type. This number of arrangements is given by the multinomial coefficient

$$\frac{(n-5)!}{4!\ 1!\,(n-10)!}.$$

(c) *Two singles and two doubles*

Of the original n circles, we lose 3 (linked in the first 3 S's or D's), plus one more for every D, giving a total of $(n-5)$ symbols. This time we have 2 S's, 2 D's and $(n-9)$ circles. Hence the number of arrangements is:

$$\frac{(n-5)!}{2!\,2!(n-9)!}$$

(d) *Three doubles*

Of the original n circles, we lose 2 for first two doubles, plus one more for every D, leaving a total of $(n-5)$ symbols. Of these, there are 3 D's and $(n-8)$ circles. So the number of arrangements is:

$$\frac{(n-5)!}{3!(n-8)!}$$

Hence the total number of arrangements is:

$$M=(n-5)!\left[\frac{1}{6!(n-11)!}+\frac{1}{4!(n-10)!}+\frac{1}{2!2!(n-9)!}+\frac{1}{3!(n-8)!}\right].$$

Finally, the probability of obtaining no runs longer than two consecutive numbers in a random selection of 6 numbers from 1, 2, ... , n is $P=\dfrac{M}{N}$, where $N=\dbinom{n}{6}$ is the number of possible selections of six numbers from 1, 2, ... , n.

For Tattslotto, $n = 45$, and hence $P = 0.9437$. Equivalently, the probability of *at least* one run of *at least* three consecutive numbers amongst the six "winning" Tattslotto numbers is $1-P=0.05627$.

A solution was also received from Carlos Victor.

PROBLEM 22.4.2 (J A Deakin, Shepparton, Vic)

Find all 2×2 matrices that commute with the matrix $\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$

SOLUTION (Julius Guest, East Bentleigh, Vic)

For commuting 2×2 matrices we here need

$$\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \qquad (1)$$

Equating corresponding elements on both sides of (1) we find

$$a + 2c = a + 4b, \text{ so } c = 2b \qquad (2)$$

and

$$4a + 3c = c + 4d \qquad (3)$$

Using (2) in (3) we obtain $a = d - b$. (4)

Equating the remaining two elements provides no further information. Hence all 2×2 matrices which commute with the quoted matrix are of type

$$\begin{bmatrix} d - b & b \\ 2b & d \end{bmatrix} \quad \text{for all } b \text{ and } d.$$

Solution also received from Carlos Victor.

PROBLEM 22.4.3 (Julius Guest, East Bentleigh, Vic)

Given that $x + y + z + u = 0$, prove that

$$x^3 + y^3 + z^3 + u^3 + 3(x+y)(y+z)(z+x) = 0$$

SOLUTION (Carlos Victor, Rio de Janeiro, Brazil)

We have $x+y+z = -u$, so that $(x+y+z)^3 = -u^3$ and then expanding the left hand side gives

$$x^3+y^3+z^3+3(xyz+x^2y+xz^2+x^2z+y^2z+xy^2+yz^2+xyz) = -u^3.$$

Now inspection of the first two terms inside the bracket gives

$$xyz+x^2y = xy(x+z),$$

and similarly the remaining three pairs also contain a factor of $x+z$ so that the above equation can be written as

$$x^3+y^3+z^3+3(x+z)(xy+xz+y^2+yz) = -u^3.$$

A similar argument applied to the term in the second bracket gives

$$x^3+y^3+z^3+3(x+z)(y+z)(x+y) = -u^3,$$

from which the desired result follows.

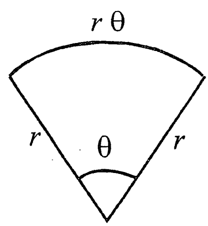
Solutions were also received from John Barton and Julius Guest.

PROBLEM 22.4.4 (from Mathematical Spectrum)

A piece of wire of length l is bent into the shape of a sector of a circle. Find the maximum area of the sector.

SOLUTION (John Barton, Carlton, Vic)

$$\begin{aligned} A &= \frac{1}{2}r^2\theta \\ &= \frac{1}{2}r^2(\tfrac{l}{r}-2) \\ &= \frac{1}{2}rl-r^2 \\ &= (\tfrac{1}{4}l)^2-(\tfrac{1}{4}l-r)^2 \end{aligned}$$



$l = r(2+\theta)$, for $0<\theta<\pi$

This last expression has a maximum value of $(\frac{1}{4}l)^2$, occurring when $r = \frac{1}{4}l$. When $r = \frac{1}{4}l$, $\theta = 2$, so that the angle of the sector is less than 2π radians.

Solutions also received from Julius Guest and Carlos Victor.

PROBLEM 22.4.5 (from the Memorial University Undergraduate Mathematics Competition, 1997)

Prove that $1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2} < 2$.

SOLUTION (David Shaw, Geelong, Vic)

By comparing terms we see that

$$1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2} < 1 + \frac{1}{2 \times 1} + \frac{1}{3 \times 2} + \ldots + \frac{1}{n(n-1)}$$

Now using the result that $\frac{1}{r(r-1)} = \frac{1}{r-1} - \frac{1}{r}$, for $r \neq 0, 1$ we see that

$$1 + \frac{1}{2 \times 1} + \frac{1}{3 \times 2} + \ldots + \frac{1}{n(n-1)} = 1 + \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \ldots + \left(\frac{1}{n-1} - \frac{1}{n}\right) = 2 - \frac{1}{n},$$

so that $1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2} < 2 - \frac{1}{n} < 2$.

Solutions were also received from Bill Tetley, Carlos Victor and John Barton.

PROBLEM 22.4.6 (from *Crux Mathematicorum with Math. Mayhem*)

In how many ways can the 12 vertices of a regular icosahedron be partitioned into four classes of three vertices, such that the vertices in each class belong to the same face?

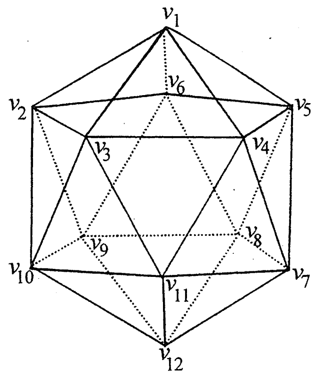
SOLUTION (Carlos Victor, Rio de Janeiro, Brazil)

In the figure below suppose that we initially choose the upper face $v_1 v_2 v_3$ as the first group of 3 vertices. From the conditions stated for the problem this choice implies that for the side faces we can have either:

(i) $v_4 v_5 v_7$ and $v_6 v_8 v_9$, or

(ii) $v_4 v_7 v_{11}$ and $v_5 v_6 v_8$

The choice (i) then implies that we must choose the lower face $v_{10} v_{11} v_{12}$ while choice (ii) gives $v_9 v_{10} v_{12}$ for the lower face. So we see that for each upper face there are two possible choices for the side face and then only one choice for the lower face. This results in $5 \times 2 \times 1 = 10$ possible choices.



PROBLEMS

Readers are invited to send in solutions (complete or partial). All solutions received by 1 May 1999 will be acknowledged in the June 1999 issue, and the best solutions will be published.

PROBLEM 22.1.1 (Adapted from *New Scientist #1597*, submitted by Greg Sheehan, Montrose, Vic)

A person has 2 pockets in their trousers and only carries amounts of money such that the sum of money in both pockets is to equal the product of the money in the right and left pockets (measured in dollars).

For example the person could carry a total of \$6.25 since it is possible to put \$5 in one pocket and \$1.25 in the other as $5+1.25 = 5\times 1.25$.

Given that the smallest unit of currency is one cent and that 100 cents equals \$1, how many different sums can be carried?

PROBLEM 22.1.2 (from *Parabola*)

In a family of 6 children, the 5 eldest children are respectively 2,6,8,12 and 14 years older than the youngest. The age of each child is a prime number.

How old are they? Show that their ages will never again be all prime numbers (even if they live indefinitely).

PROBLEM 22.1.3 (Julius Guest, East Bentleigh, Vic)

Solve the system

$$\begin{aligned} x+y+z &= 9 \\ x^2+y^2+z^2 &= 29 \\ x^3+y^3+z^3 &= 99 \end{aligned}$$

PROBLEM 22.1.4 (K.R.S. Sastry, Bangalore, India)

Let P be a point inside a square $ABCD$.

Prove that P lies on the diagonal AC if and only if PA^2, PB^2, PC^2 are in arithmetical progression, in that order.

PROBLEM 22.1.5 (from *Crux Mathematicorum with Math. Mayhem*)

Suppose that a, b, c are positive real numbers such that

$$abc = (a+b-c)(b+c-a)(c+a-b).$$

Clearly $a = b = c$ is a solution. Determine all others.

* * * * *

Published by Department of Mathematics & Statistics, Monash University