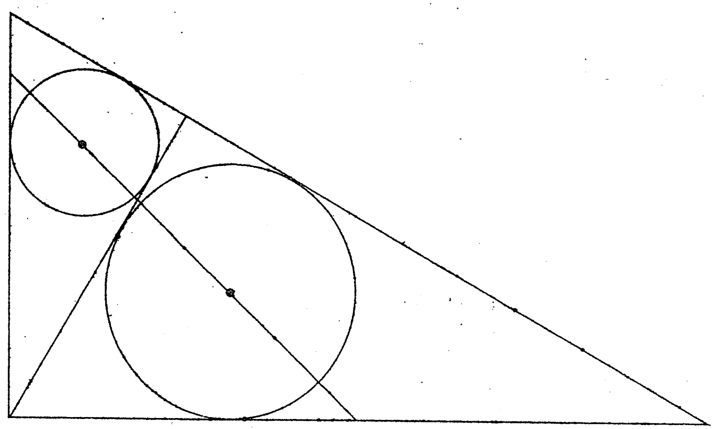
FUNCTION

Volume 12 Part 5 October 1988



A SCHOOL MATHEMATICS MAGAZINE
Published by Monash University

Reg. by Aust Post Publ. No. VBH01[illegible]

FUNCTION is a mathematics magazine addressed principally to students in the upper forms of schools, and published by Monash University.

It is a 'special interest' journal for those who are interested in mathematics. Windsurfers, chess-players and gardeners all have magazines that cater to their interests. FUNCTION is a counterpart of these.

Coverage is wide - pure mathematics, statistics, computer science and applications of mathematics are all included. There are articles on recent advances in mathematics, news items on mathematics and its applications, special interest matters, such as computer chess, problems and solutions, discussions, cover diagrams, even cartoons.

* * * * *

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 mathematics departments of the institutions shown above.

The magazine is published five times a year, appearing in February, April, June, August, October. Price for five issues (including postage): \$10.00*; single issues \$2.00. 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.

* \$5.00 for *bona fide* secondary or tertiary students.

* * * * *

Registered for posting as a periodical - "Category B"
ISSN 0313 - 6825

As announced in the last issue of *FUNCTION* Terry Tao of Blackwood High School, South Australia, obtained a gold medal in this year's International Mathematical Olympiad, held this year in Canberra, and one of Australia's bicentennial events. Terry's solution to one of the questions, question 5, one of the three questions set on the second day of the competition, is included (as Solution 8) in the article by Emanuel Strzelecki, Australia's team leader. Dr Strzelecki provides eight solutions, some of which have considerable differences from each other.

The *cover diagram* is to illustrate the problem: show that the area of the largest triangle is greater than or equal to twice the area of the other triangle with which it shares a common right angle.

CONTENTS

QUESTIONS 5 AND 6 OF THE INTERNATIONAL MATHEMATICAL OLYMPIAD 1988

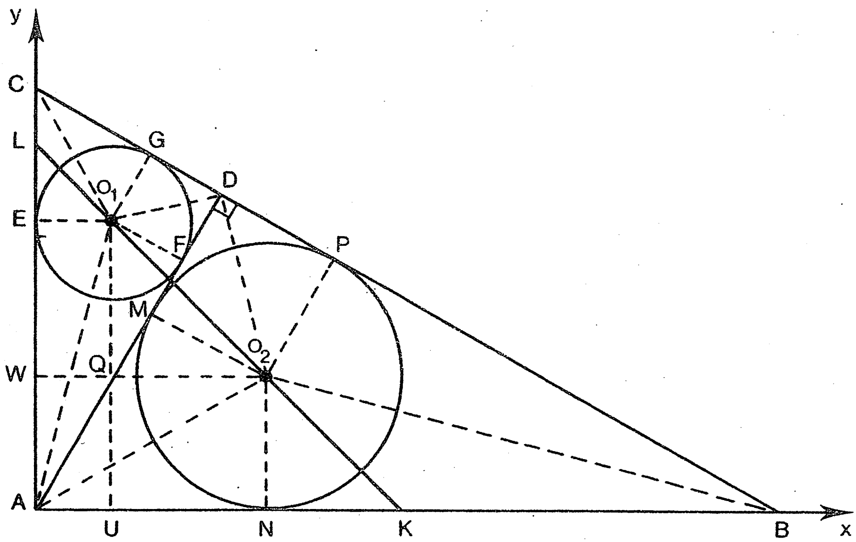
Emanuel Strzelecki
Team Leader, Australian team

This year the International Mathematical Olympiad was conducted in Canberra, Australia.

On the second day of the competition two very interesting problems were offered. First we discuss a number of solutions for question 5.

QUESTION 5. ABC is a triangle right-angled at A, and D is the foot of the altitude from A. The straight line joining the incentres of the triangles ABD, ACD intersects the sides AB, AC at the points K, L respectively. S and T denote the areas of the triangles ABC and AKL respectively. Show that $S \geq 2T$.

The figure below has marked on it all the subsidiary lines and points necessary to cover all the solutions we offer.



Let us consider first some theorems that will be useful in solving the problem.

Theorem 1

If ABC is a triangle right-angled at A, and D is the foot of the altitude from A then triangles DAC, DBA and ABC are similar, since

$$C\hat{A}D = C\hat{B}A, \quad D\hat{A}B = A\hat{C}B.$$

Theorem 2

In similar triangles all corresponding segments (for example sides, altitudes, radii of incircles, etc.) are in the same ratio.

The crucial point in solving the problem is to show that $AL = AK = AD$ (Fig.). Once the above equality has been established we can proceed, for example, as follows:

$$T = \frac{1}{2}AL \times AK = \frac{1}{2}(AD)^2.$$

Since in a right-angled triangle

$$AD = \sqrt{CD \times DB},$$

and, by the geometric mean, arithmetic mean inequality

$$\sqrt{CD \times DB} \le \frac{1}{2}(CD+DB) = \frac{1}{2}BC,$$

we obtain

$$T = \frac{1}{2}AD \times AD \le \frac{1}{2}AD \times \frac{1}{2}BC = \frac{1}{2}S,$$

as required.

We can also use trigonometry:

From $\triangle$ ADC and $\triangle$ ABC

$$AD = AC \sin C = BC \sin C \cos C.$$

So

$$T = \frac{1}{2}AD \times AD = \frac{1}{2}AD \times BC \sin C \cos C$$

$$= S \times \frac{\sin 2C}{2} \le \frac{1}{2}S.$$

Below we present a few proofs of the equality

$$AL = AK = AD.$$

Solution 1

Since $B\hat{A}C = 90°$ we can introduce a system of coordinates with A as the origin and AB and AC as the x- and y-axes, respectively. Denote by E, F, G the points of contact of the circle inscribed in the triangle ADC with its sides, and by M, N, P the points of contact of the circle inscribed in the triangle ADB with its sides. Let $O_1F = r$, $O_2M = R$. Let $O_1U \perp AB$ and $O_2W \perp AC$.

Since $O_1E \perp AC$, O_1EAU is a rectangle, and thus

$$O_1U = EA. \tag{1}$$

By the properties of tangents to a circle from a common point, we have

$$AE = AF. \tag{2}$$

Since $O_1G \perp DC$, $O_1F \perp AD$ and $O_1G = O_1F = r$, O_1FDG is a square, and thus

$$AF = AD - FD = AD - r. \tag{3}$$

From (1), (2) and (3) we obtain

$$O_1U = AD - r.$$

Similarly

$$O_2W = AN = AM = AD - R.$$

Thus the gradient of O_1O_2 is

$$m_{KL} = \frac{O_1U-O_2N}{O_1E-O_2W} = \frac{AD-r-R}{r-AD+R} = -1.$$

It follows that $L\hat{K}A = 45°$ and hence

$$AK = AN + NK = AN + O_2N = AM + R = AD.$$

Similarly $AL = AD$.

Solution 2 (Variation of a solution offered by John M. Mack, University of Sydney)

Let $AL = AK = AD$ (Fig.). Let AO_2 be the bisector of the angle BAD, with O_2 the intersection of the bisector with KL. Join O_2 with D.

Then, since
1) $AK = AD$
2) $AO_2 = AO_2$ (common)
3) $K\hat{A}O_2 = D\hat{A}O_2$,

we have

$$\Delta\ AKO_2 = \Delta\ ADO_2.$$

But $L\hat{K}A = 45°$, since $AK = AL$ and $L\hat{A}K = 90°$.
So $A\hat{D}O_2 = 45° = O_2\hat{D}P$.

So O_2 is the intersection of bisectors of two angles of the triangle ABD. Hence O_2 is the incentre of the triangle ABD.

It follows that the incentre of the Δ ABD lies on KL. Similarly we prove that the incentre of the triangle ADC is on KL.

This proves the required equality.

Solution 3

Let O_1 and O_2 be the incentres of triangles ADC and ADB respectively. Consider the triangle O_1DO_2 (Fig. 1). We have

$$O_1\hat{D}O_2 = O_1\hat{D}A + A\hat{D}O_2 = 45° + 45° = 90°.$$

Also, since triangles ADC and ADB are similar, we have

$$O_1D/O_2D = AC/AB$$

(In similar triangles any corresponding segments are proportional.)

It follows that the triangle O_1O_2D is similar to the triangle ADC. Hence triangle O_1DO_2 can be considered as obtained by rotating the triangle ADB around the point D through an angle $O_1\hat{D}A = 45°$ in the clockwise direction and reducing it in size in the ratio O_1D to AD.

It follows that O_1O_2 makes an angle of $45°$ with AB.

The rest is as in Solution 1.

Solution 4

The following version of Solution 3 was offered by Angelo Di Pasquale, a year 9 student from Eltham College, who attended Friday lectures on mathematical problems at Presbyterian Ladies' College (PLC).

After it has been established that $\Delta\ O_1O_2D \approx \Delta$ ABD, we conclude that

$O_1O_2D = A\hat{B}D$. It follows that in the quadrilateral $KBDO_2$

$$D\hat{O}_2K + K\hat{B}D = 180° - O_1\hat{O}_2D + A\hat{B}D = 180°.$$

So the quadrilateral is cyclic, and hence

$$O_2\hat{K}B + O_2\hat{D}B = 180°.$$

Thus $O_2\hat{K}A = 180° - O_2\hat{K}B = O_2\hat{D}B = 45°$.
The rest is as in Solution 2.

Solution 5

From trapezium O_2O_1GP we have

$$\tan O_2\hat{O}_1F = \frac{O_2P-O_1G}{PG} = \frac{R-r}{PD+DG} = \frac{R-r}{R+r}.$$

Denote AC by b, AB by c. Since triangles ADC and ABD are similar, we have

$$\tan O_2\hat{O}_1F = \frac{R-r}{R+r} = \frac{c-b}{c+b}.$$

Since $O_1F \parallel BC$ and $O_1U \parallel AC$ we have $F\hat{O}_1U = A\hat{C}B$.

Thus $O_2\hat{O}_1U = F\hat{O}_1U - F\hat{O}_1O_2 = A\hat{C}B - F\hat{O}_1O_2$. Hence

$$\tan O_2\hat{O}_1U = \frac{\tan A\hat{C}B - \tan F\hat{O}_1O_2}{1+\tan A\hat{C}B \times \tan F\hat{O}_1O_2}$$

$$= \frac{\frac{c}{b} - \frac{c-b}{c+b}}{1+\frac{c}{b}\times\frac{c-b}{c+b}} = \frac{c^2+b^2}{b^2+c^2} = 1 ;$$

consequently, $O_2\hat{O}_1U = 45°$.

The rest is as in Solution 1.

Solution 6

Let $O_2W \perp AC$, $O_1U \perp AB$ and let Q be the intersection of O_2W and O_1U. Then

$$\tan Q\hat{A}B = \frac{QU}{QW} = \frac{R}{r} = \frac{c}{b} = \tan A\hat{C}B.$$

So $Q\hat{A}B = A\hat{C}B$. Also $D\hat{A}B = A\hat{C}B$. It follows that Q is on AD.

Thus $Q\hat{O_2}A = O_2\hat{A}N = O_2\hat{A}Q$. Hence $O_2Q = AQ$. Similarly $O_1Q = AQ$. Consequently $O_1Q = O_2Q$, and since $O_1\hat{Q}O_2 = 90°$ we obtain

$$O_1\hat{O_2}Q = 45°.$$

The rest is as in Solution 1.

Solution 7

Note that the triangle ADC can be obtained from the triangle BDA by rotation around D through an angle of 90° in the clockwise direction, followed by a change of size in the ratio of AC to AB.

It follows that all corresponding segments in the above triangles are perpendicular to each other. In particular,

$$BO_2 \perp AO_1 \quad \text{and} \quad AO_2 \perp CO_1.$$

Thus bisectors CO_1 and BO_2 of the angles ACB and ABC of the triangle ABC when extended are altitudes of the triangle AO_1O_2. Since both the bisector of the angle BAC and the altitude of triangle O_1AO_2 from the vertex A contain A and the point of intersection of bisectors CO_1 and BO_2, they coincide.

It follows that O_1O_2 is perpendicular to the bisector of the angle BAC and hence $L\hat{K}M = 45°$.

The rest is as in Solution 1.

Solution 8. (Terry Tao, gold medalist in the Australian team, IMO 1988)

Let $A\hat{B}C = \beta$, $A\hat{C}B = \gamma$. Then $O_2\hat{B}D = \frac{\beta}{2}$, $O_2\hat{D}B = 45°$, and $B\hat{A}D = A\hat{C}B = \gamma$. Applying the sine rule to the triangle O_2DB we obtain

$$O_2D = \frac{BD \cdot \sin\frac{\beta}{2}}{\sin(45°+\frac{\beta}{2})}.$$

From $\triangle$ ADB, $BD = c \sin\gamma$.

$$\text{So} \quad O_2D = \frac{c \sin\gamma \cdot \sin\frac{\beta}{2}}{\sin(45°+\frac{\beta}{2})} = \frac{c \cos\beta \cdot \sin\frac{\beta}{2}}{\sin(45°+\frac{\beta}{2})}.$$

Similarly

$$O_1D = \frac{b \sin \beta \cdot \sin \frac{\gamma}{2}}{\sin(45^\circ+\frac{\gamma}{2})} = \frac{b \sin \beta \cdot \sin \frac{90^\circ-\beta}{2}}{\cos(45^\circ-\frac{\gamma}{2})} = \frac{b \sin \beta \sin(45^\circ-\frac{\beta}{2})}{\cos \frac{\beta}{2}}.$$

Since $O_1\hat{D}O_2 = 90^\circ$, from $\triangle O_1DO_2$ we have

$$\tan O_2\hat{O}_1D = \frac{O_2D}{O_1D} = \frac{c}{b} \frac{2 \cos \beta \sin \frac{\beta}{2} \cos \frac{\beta}{2}}{2 \cos(45^\circ-\frac{\beta}{2}) \cdot \sin \beta \sin(45^\circ-\frac{\beta}{2})}$$

$$= \frac{c}{b} \frac{\cos \beta \sin \beta}{\sin \beta \sin(90^\circ-\beta)} = \frac{c}{b} = \tan \gamma.$$

Hence $O_2\hat{O}_1D = \gamma$. Thus $O_1\hat{O}_2D = \beta$.

Consequently $K\hat{O}_2B = 180^\circ - O_1\hat{O}_2D - D\hat{O}_2B$

$$= 180^\circ - \beta - (180^\circ\text{-}45^\circ\text{-}\tfrac{\beta}{2}) = 45^\circ - \tfrac{\beta}{2}.$$

Thus $O_2\hat{K}A = K\hat{O}_2B + O_2\hat{B}K = 45^\circ - \frac{\beta}{2} + \frac{\beta}{2} = 45^\circ$.

The rest is as in previous solutions.

We now turn to question 6.

QUESTION 6. Let a and b be positive integers such that $ab + 1$ divides $a^2 + b^2$. Show that $\frac{a^2+b^2}{ab+1}$ is the square of an integer.

Question 6 was found the most difficult of all questions this year. Only 11 of the 268 contestants got the full 7 marks and only 14 got 4 marks or more. For question 5, 86 contestants got full marks and 124 got 5 marks or more.

To solve question 6, consider the equation

$$\frac{x^2+y^2}{xy+1} = q \tag{1}$$

where q is a positive integer. This equation has the same solutions x, y as

$$x^2 + y^2 = q(xy+1); \tag{2}$$

for the only extra solutions that could occur would be when $xy + 1 = 0$ (when (1) would not make sense), and this is impossible, for then, from (2), $x = y = 0$.

Thus it suffices to study equation (2). All 11 complete solutions essentially used the following idea. Suppose that $x = b,\ y = a$ is an integer solution of (2). Then $x = b$ is a solution of the quadratic equation

$$x^2 + a^2 = q(ax+1) \qquad (3)$$

i.e.

$$x^2 - qax + a^2 - q = 0. \qquad (4)$$

Let $x = c$ be the other solution of (4). Then (4) must be the same equation as

$$(x-b)(x-c) = 0$$

i.e.

$$x^2 - (b+c)x + bc = 0. \qquad (5)$$

Comparison of the (identical) equations (4) and (5) shows that $b + c = qa$; so, since q, a and b are integers, c is also an integer.

Thus if we start with one pair of solutions b, a of (2) we can construct another pair c, a, unless of course $c = b$.

Now let us assume that q is not the square of an integer, and that (2) has positive integer solutions. Choose a solution $x = b,\ y = a$ such that a is the least number occurring in any solution pair.

Then, as we have seen, there is a second pair $x = c,\ y = a$, also a solution of (2). We claim that $0 < c < a$; and this contradicts the choice of a as the least integer occurring in a solution. This contradiction shows that there can be no solutions when q is not the square of an integer.

Now to prove that $0 < c < a$. First note that, if c is a negative integer, then from (3) we have the contradiction between

$$c^2 + a^2 > 0 \quad \text{and} \quad q(ca+1) < 0.$$

Hence $c \geq 0$.

If $c = 0$, then from (5) and (4)

$$0 = bc = a^2 - q,$$

so that q is the square of an integer, contrary to assumption.

If $c \geq a$, then again, from $bc = a^2 - q$, since also $b \geq a$, we have $a^2 - q \geq a^2$, which is impossible. Hence

$$0 < c < a$$

which we have also already seen to be impossible.

This concludes the solution to question 6.

Problem 12.5.1 Find all positive integer pairs of solutions x, y of

$$\frac{x^2+y^2}{xy+1} = k^2,$$

where k is a positive integer.

29th International Mathematical Olympiad 1988

First Day, Canberra, July 15

Time allowed: 4.5 hours

1. Consider two coplanar circles of radii R and r $(R > r)$ with the same centre. Let P be a fixed point on the smaller circle and B a variable point on the larger circle. The line BP meets the larger circle again at C. The perpendicular ℓ to BP at P meets the smaller circle again at A (if ℓ is tangent to the circle at P then $A = P$).
   (i) Find the set of values of $BC^2 + CA^2 + AB^2$.
   (ii) Find the locus of the midpoint of AB.

2. Let n be a positive integer and let $A_1, A_2, \ldots, A_{2n+1}$ be subsets of a set B. Suppose that
   (a) each A_i has exactly $2n$ elements,
   (b) each $A_i \cap A_j$ $(1 \le i < j \le 2n + 1)$ contains exactly one element, and
   (c) every element of B belongs to at least two of the A_i.

   For which values of n can one assign to every element of B one of the numbers 0 and 1 in such a way that each A_i has 0 assigned to exactly n of its elements?

3. A function f is defined on the positive integers by

$$f(1) = 1,\ f(3) = 3,$$
$$f(2n) = f(n),$$
$$f(4n + 1) = 2f(2n + 1) - f(n),$$
$$f(4n + 3) = 3f(2n + 1) - 2f(n),$$

   for all positive integers n.

   Determine the number of positive integers n, less than or equal to 1988, for which $f(n) = n$.

Questions continued on p. 144.

KNOTS[†]

John Stillwell
Department of Mathematics, Monash University

One of the interesting, though sometimes frustrating, things about mathematics is that "obvious" facts can be hard to prove. For example, it is surely obvious that the *trefoil curve* (shown in Fig. 1) is knotted.

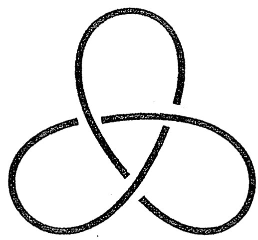


Fig. 1

So obvious, in fact, that most people would not consider this to be a mathematical fact at all. But what *does* it mean for a curve to be knotted, and how could knottedness of a given curve be rigorously proved?

The most elementary way to do this is to model curves by polygons without self-intersections in 3-dimensional space. The trefoil could be taken in the polygonal form shown in Fig. 2.

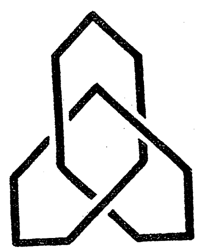


Fig. 2.

[†] This article originally appeared in John Stillwell's "World of Mathematics" column in *Vinculum* (published by the Mathematical Association of Victoria).

Alternatively, we could imagine that the smooth picture originally given for the trefoil was actually a polygon with a very large number of tiny sides.

The point of viewing knots in this finite, polygonal way is that all possible deformations of the curve can be broken down into simple atomic moves, shown in Fig. 3. (Fig. 3 shows that happens to one edge of the polygon. The whole polygon is a closed curve – which is just as well, since any curve with loose ends can be untied.)

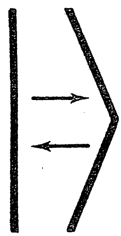


Fig. 3.

Each move is the "bending" of one edge into two, or the reverse. Of course, such a move is allowed only if it does not cause the polygon to pass through itself. Sequences of atomic moves can obviously model any actual deformation of a physical curve (made of string or whatever) with arbitrary precision.

With this model of curves and their deformations, we can give a precise definition of what it means for a curve K to be knotted: *K is knotted if there is no sequence of atomic moves which convert K to a standard unknotted polygon, say a triangle.* (Thus we have really said that a knotted curve is one which is not *un*knotted. The property of being unknotted – i.e. convertible to a triangle by basic moves – is easy to understand.) The problem of deciding whether a curve is knotted can now be regarded as a mathematical problem, but still a rather hard one. Even the trefoil has never been proved to be knotted directly from first principles – all proofs require additional geometric or algebraic machinery.

The simplest proof I know of was given recently by Lou Kauffmann in the *American Mathematical Monthly*, vol. 95 (1988), p. 203. The proof is based on two old ideas in knot theory : *knot projections* and *Reidemeister moves.*

Knot projections are simply planar pictures of curves – like those we have already drawn – in which no more than two lines cross at any point. The convention is that the lower line is drawn broken at a crossing point, so that the picture consists of a series of arcs. It is obvious (*and* easy to prove rigorously) that any polygon K can be deformed by atomic moves so as to have such a projection.

Reidemeister moves are the basic nontrivial changes which can occur in the projection of a polygon K as a result of atomic moves on K. They are of three types, I, II, III, shown in Fig. 4

I: II: III:

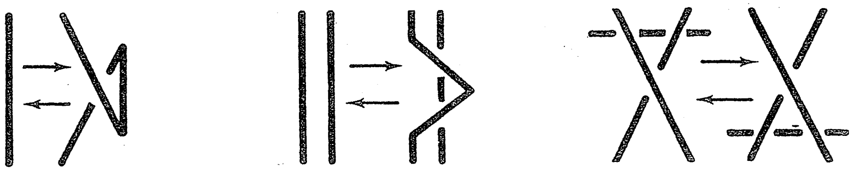


Fig. 4

This gives us another definition of knottedness, which turns out to be easier to work with: K *is knotted if there is no sequence of Reidemeister moves which convert a projection of K to a simple polygon (one without crossings).* To use this definition to prove knottedness of some particular K, one wants to find a property of K which is preserved by Reidemeister moves and which is *not* a property of simple polygons.

We shall give such a property for the trefoil in a moment, but first it may help to consider a simpler property which can be used to demonstrate not knotting, but *linking* of curves.

Two polygons A, B are said to be *linked* if their projection cannot be converted, by Reidemeister moves, to one in which A has no crossings with B. The "obvious" example of linked curves is of course Fig. 5.

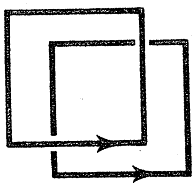
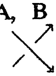


Fig. 5.

We define the *linking number* of directed curves A, B as the sum of numbers ± 1 at crossings of A with B. A crossing  has value + 1 (the '+' indicates that you move anticlockwise in going from the upper portion of the directed curve to that portion lying underneath at the crossing), while  has value - 1. The linking number is tailor-made to be unchanged by Reidemeister moves: moves I and III do not change the crossings of A with B at all (though III may move one of them), while move II at worst introduces or removes two crossings of opposite sign. Also, changing the direction of A or B can only change the sign of the linking number. Since unlinked curves have linking number 0, it follows that the property of *having nonzero linking number* is one which is preserved by Reidemeister moves and which is *not* a property of unlinked polygons.

Inspection shows that the polygons in Fig. 5 have linking number 2, so this proves that they are linked.

There seems to be no property quite as simple as this which guarantees knotting, but in the case of the trefoil we can come close.

The trefoil has a property which Kauffmann calls tricolorability, illustrated in Fig. 6.

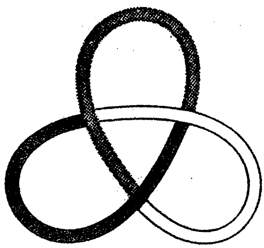


Fig. 6

A projection is said to be tricolorable if its arcs can each be given one of the three colours (so colours change only where the curve passes under another portion of the curve), in such a way that all three colours are used and one or three colours occur at each crossing. To make life easier for the printer I have made the three colours white, grey and black, and to show them more plainly I have fattened the edges of the projection. A simple polygon is *not* tricolorable because it has only one arc, so only one colour can be used.

Thus to prove that the trefoil is knotted it will suffice to check that tricolorability is preserved by Reidemeister moves. For I moves, the colouring need not be changed at all. For a II move in the forward direction, involving different coloured strands, we have to introduce a new colour as in Fig. 7.

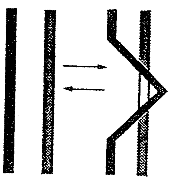


Fig. 7.

Conversely, if we have a II move in the backward direction, involving different coloured strands, then the colours have to be distributed as in the right half of Fig. 7, in which case we can move back to the colouring in the left half. Finally it can be checked that the only multi-colourings which can occur in the context of III moves are those shown in Fig. 8, which also shows that we can move from one colouring to the other.

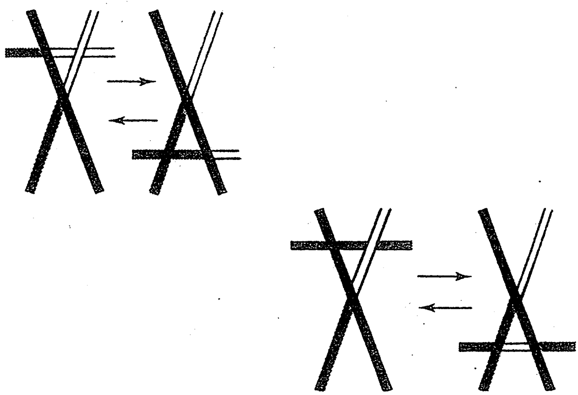


Fig. 8.

So now we have proved that the trefoil is knotted!

This very elementary method works for quite a few knots – about 20% of the knots with ≤ 9 crossings are tricolourable according to my experiments with coloured pencils. To prove that other curves are knotted, more sophisticated methods have been brought into play. In fact, the methods used to date in knot theory include much of the geometry, algebra and topology developed over the last century.

However, an encouraging trend in recent research has been the resurgence of quite elementary ideas, such as the Reidemeister moves. This new era in knot theory dawned in 1985 with the discovery of a new knot invariant called the *Jones polynomial*. Its discoverer, Vaughn Jones, was not even working in knot theory, but in the esoteric field of von Neumann algebras, when he discovered some formulas which led in a mysterious way to knots. The connection still remains mysterious – it seems to imply quite extraordinary connections between knots and physics – but its practical consequence is a polynomial V_K which can be computed for each polygon K, and which is unchanged by Reidemeister moves. $V_K = 1$ for a simple polygon K, hence $V_K \neq 1$ guarantees that K is knotted.

Such polynomials had been known since the 1920's, but the Jones polynomial outperforms them in most cases. Not only does it provide a proof of knottedness for all known knots, it also shows that certain knots are distinct from their mirror images. This is another "obvious" property of the trefoil which until now had been hard to prove.

How easy is it now? As I said, Jones arrived at his polynomial via a very roundabout route. Fortunately, Kauffmann has since found a simple explanation of the Jones polynomial. He relates it to knot projections in such a way that it becomes easy to see why the polynomial is unchanged by Reidemeister moves. If you followed my explanation of tricolorability, you may care to read Kauffmann's explanation of the Jones polynomial which is in the article cited above.

The main open problem concerning the Jones polynomial is whether it recognizes knottedness in all cases. That is, do knotted polygons K invariably have $V_K \neq 1$? If so, this would be the first tolerably simple method for recognising knottedness. If not, then it's back to the drawing board. Perhaps some generalisation of crossing number and/or tricolorability will work? This is an area in which a young mathematician might easily get results.

* * * * *

29th IMO 1988 (Continued)

Second Day, Canberra, July 16

Time allowed: 4.5 hours

4. Show that the set of real numbers x which satisfy the inequality

$$\sum_{k=1}^{70} \frac{k}{x-k} \geq \frac{5}{4}$$

is a union of disjoint intervals, the sum of whose lengths is 1988.

5. ABC is a triangle right-angled at A, and D is the foot of the altitude from A. The straight line joining the incentres of the triangles ABD, ACD intersects the sides AB, AC at the points K, L respectively. S and T denote the areas of the triangles ABC and AKL respectively. Show that $S \geq 2T$.

6. Let a and b be positive integers such that $ab + 1$ divides $a^2 + b^2$. Show that $\frac{a^2+b^2}{ab+1}$ is the square of an integer.

THE RULE OF 72

Michael A.B. Deakin, Monash University

If money is invested or borrowed at compound interest, the investment or debt grows exponentially as time goes by. After n years, for each initial dollar we now have

$$f(n) = (1 + \frac{r}{100})^n \tag{1}$$

dollars, where r is the interest rate expressed as a percentage.

If we wait long enough, each initial dollar will become two dollars – the initial investment or debt will have doubled. Let T be the time required for each dollar to double. Then

$$(1 + \frac{r}{100})^T = 2. \tag{2}$$

Recently, I learned† of a simple rule for estimating T; a friend tells me he learned of it from the financial pages of one of the daily papers. It is called "the rule of 72" and it reads

$$T \simeq 72/r. \tag{3}$$

Let us investigate how good this approximation is. Take logs of both sides of Equation (2). We may do this to any base at all, but it will be convenient to do this to base e and use so-called natural logarithms. Most calculators have these in convenient form. Using the notation adopted on the keyboards of such calculators, we may write Equation (2) as

$$T \cdot \ln(1 + \frac{r}{100}) = \ln 2 \simeq 0.693 \ldots$$

or

$$\begin{aligned} T &= \frac{\ln 2}{\ln(1+\frac{r}{100})} \\ &= \frac{r \ln 2}{\ln(1+\frac{r}{100})} \Big/ r \\ &= \phi(r)/r \qquad \text{(say)}. \end{aligned}$$

Thus we have, and this is exact,

$$T = \phi(r)/r. \tag{4}$$

---

† See pp. 62-3 of *Making Money Made Simple*, by Noel Whittaker, Boolarong Publications, Brisbane, 1988.

When r is small

$$\ln(1 + \frac{r}{100}) \approx \frac{r}{100}, \tag{5}$$

the validity of this formula depending on our choice of *natural* logs and being the reason for this choice. Thus for small r

$$\phi(r) \approx 100 \ln 2 \approx 69.3.$$

As r increases, so does $\phi(r)$, and we can easily calculate the entries in the table below:

r	$\phi(r)$
0	69.3
5	71.0
10	72.7
15	74.4
20	76.0
25	77.7

Table 1.

Thus $\phi(r) \approx 72$ for a value of r near 10 and Approximation (3) works best for interest rates of about 10%. Above this value it tends to underestimate T to some extent (although the *proportional* error increases). However, as Table 2 shows, it gives reasonably good results over a wide range of values of r.

r	T	72/r
0	∞	∞
1	69.7	72.0
2	35.0	36.0
5	14.2	14.0
10	7.3	7.2
15	5.0	4.8
20	3.8	3.6
25	3.1	2.9

Table 2.

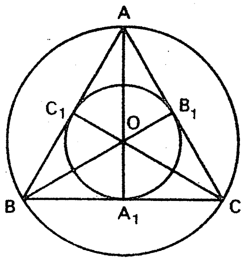
* * * * *

THE THEOREMS OF CEVA AND MENELAUS

Marta Sved[†]

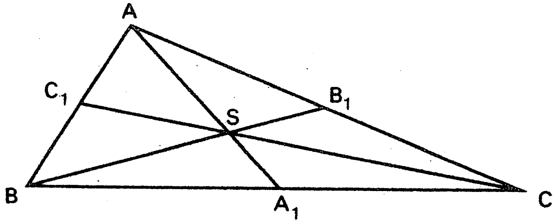
University of Adelaide

One of the simplest plane figures appealing by its symmetry is the equilateral triangle.



The figure shows its lines of symmetry AA_1, BB_1, CC_1. They meet in O, the centre of the triangle. The symmetry here is perfect. O is equidistant from the vertices A, B, C, hence it is the centre of the circumcircle of the triangle; it is equidistant from the sides BC, CA, AB, hence it is the centre of the inscribed circle. If the triangle is a physical object, i.e. a thin, uniform plate, then O is its centre of mass (or centre of gravity). The symmetry lines bisect the sides BC, CA, AB at A_1, B_1, C_1 respectively. They are perpendicular to those lines and bisect the angles at A, B and C.

Consider now what happens if the equilateral triangle is pulled out of shape to form a general (scalene) triangle.

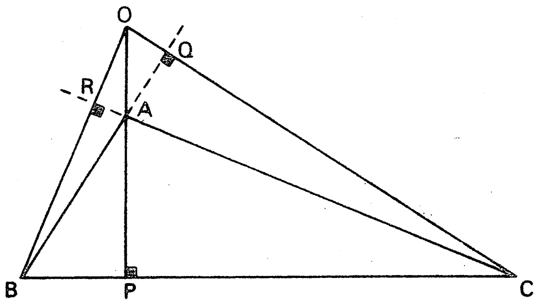


We cannot speak any more about lines of symmetry, though we can draw the medians AA_1, BB_1, CC_1, where A_1, B_1, C_1 are midpoints of the corresponding

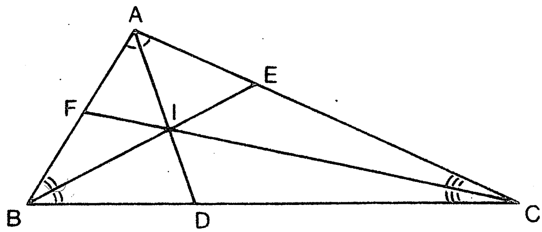
---

[†] Problem 12.4.1 (p. 127 of the previous issue of *Function*, Vol. 12, part 4) asks for a test for concurrency related to Ceva's theorem.

sides. These lines meet in one point S, called the centroid of the triangle, or if the triangle is a physical, uniform, thin plate, then S is its centre of mass. However, the lines AA_1, BB_1, CC_1 are generally not perpendicular to the sides and do not bisect the angles.



However, we may draw the altitudes AP, BQ, CR of the triangle, these being perpendicular to the sides BC, CA, AB respectively. These lines again concur at one point : O, called the orthocentre.



We may also draw the lines AD, BE, CF which bisect the angles at A, B and C. These lines also intersect in one point, I, which is the centre of the inscribed circle of the triangle.

You may have had in class the proofs of these three concurrence theorems, or you may read them in some geometry book, or may find proofs by your own efforts. In this article we want to discuss a theorem which covers all three of the above situations and many others. To establish the concurrence of three lines without using coordinate geometry (which may in some cases involve heavy algebra) can be quite difficult. The theorem to be stated and proved in the following provides a very useful tool.

CEVA'S THEOREM

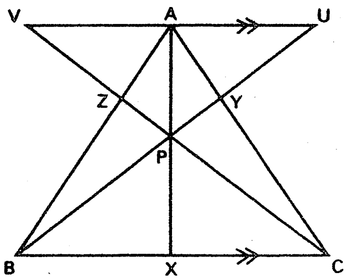
Let ABC be a triangle and let X,Y,Z lie on the segments BC,CA,AB respectively. The lines AX,BY,CZ are concurrent if and only if

$$\frac{\overrightarrow{AZ}}{\overrightarrow{ZB}} \cdot \frac{\overrightarrow{BX}}{\overrightarrow{XC}} \cdot \frac{\overrightarrow{CY}}{\overrightarrow{YA}} = 1,$$

where $\overrightarrow{AZ}$, $\overrightarrow{ZB}$, etc. denote directed[†] *lengths.*

Proof.

Suppose first that the lines AX, BY, CZ concur at P. Through A draw a line parallel to BC and produce CZ and BY to meet this line at V and U respectively. Then the triangles PAV and PXC are equi-angular and so



$$\left|\frac{\overrightarrow{XC}}{\overrightarrow{AV}}\right| = \left|\frac{\overrightarrow{PX}}{\overrightarrow{AP}}\right| .$$

Note that so far we have not paid attention to the signs of the line-segments involved and considered only absolute[†] values. We have similarly

$$\left|\frac{\overrightarrow{BX}}{\overrightarrow{AU}}\right| = \left|\frac{\overrightarrow{PX}}{\overrightarrow{AP}}\right| .$$

and so

$$\left|\frac{\overrightarrow{BX}}{\overrightarrow{XC}}\right| = \left|\frac{\overrightarrow{AU}}{\overrightarrow{AV}}\right| . \tag{1}$$

Next we consider the equi-angular triangles YBC and YUA and obtain

$$\left|\frac{\overrightarrow{CY}}{\overrightarrow{YA}}\right| = \left|\frac{\overrightarrow{BC}}{\overrightarrow{AU}}\right| . \tag{2}$$

[†] *Directed lengths*: give a line a direction and take two points P, Q on the line. Then $\overrightarrow{PQ}$ denotes the value of the length of PQ if, when moving from P to Q on the line, you go in the given direction of the line and it denotes minus the value of the length of PQ if going from P to Q is in the direction opposite to the given direction on the line. In either case $|\overrightarrow{PQ}|$, called the *absolute value*, or equivalently the *modulus*, of $\overrightarrow{PQ}$, denotes the value of the length of $\overrightarrow{PQ}$ and is thus a non-negative number.

Finally, comparing triangles ZBC and ZAV we have

$$\left|\frac{\vec{AZ}}{\vec{ZB}}\right| = \left|\frac{\vec{AV}}{\vec{BC}}\right| . \tag{3}$$

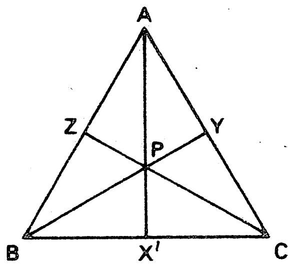
Multiplying together the left-hand sides of (1), (2), (3) we obtain

$$\left|\frac{\vec{AZ}}{\vec{ZB}} \circ \frac{\vec{BX}}{\vec{XC}} \circ \frac{\vec{CY}}{\vec{YA}}\right|$$

while the product of the right-hand sides is 1. We note now that the expression inside the modulus† sign must be positive, since the directions of AZ and ZB are the same. Similarly the other factors are positive. Hence we have proved that for the lines AX, BY, CZ to be concurrent it is *necessary* that

$$\frac{\vec{AZ}}{\vec{ZB}} \circ \frac{\vec{BX}}{\vec{XC}} \circ \frac{\vec{CY}}{\vec{YA}} = 1. \tag{4}$$

Next we prove the *converse*, that is, that if (4) holds then the lines are concurrent. Denote now the intersection of BY and CZ by P and join AP and produce it to meet BC in X'. Then for the three concurrent lines AX',BY,CZ we must have



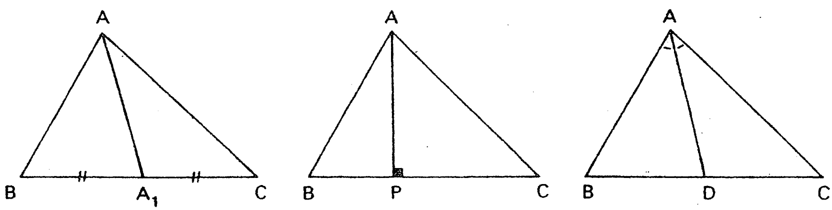
$$\frac{\vec{AZ}}{\vec{ZB}} \circ \frac{\vec{BX'}}{\vec{X'C}} \circ \frac{\vec{CY}}{\vec{YA}} = 1.$$

Comparing this with (4) we obtain that

$$\frac{\vec{BX}}{\vec{XC}} = \frac{\vec{BX'}}{\vec{X'C}} .$$

This happens only if $X' = X$, that is, if the line AP intersects BC in X. This proves that (4) is a *sufficient* condition for the concurrence and thus the proof is complete.

Applications: Consider the cases of the centroid, the orthocentre and incentre.



We find the ratios of the segments on the side BC in each case, obtaining

$$\frac{\overrightarrow{BA_1}}{\overrightarrow{A_1C}} = 1, \quad \frac{\overrightarrow{BP}}{\overrightarrow{PC}} = \frac{\tan B\hat{A}P}{\tan P\hat{A}C}, \quad \frac{\overrightarrow{BD}}{\overrightarrow{DC}} = \frac{|AB|}{|AC|}$$

[You may not be familiar with the last relation. Try to look it up or prove it by drawing through C a line parallel to AD to intersect BA produced in some point Q.]

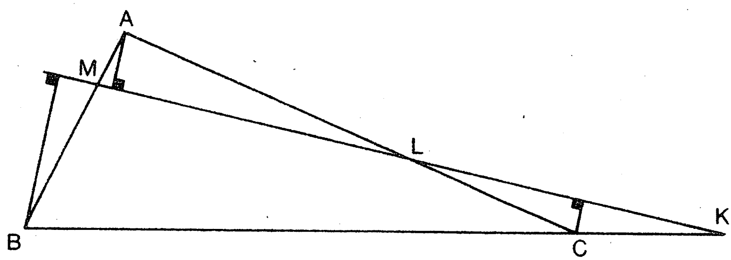
Finding the corresponding ratios for the other sides, Ceva's theorem proves the concurrence for each case. You may fill in the details for yourself.

A theorem closely allied to Ceva's theorem deals with the collinearity of three points on the sides of a triangle. We state this theorem below.

MENELAUS' THEOREM

Let ABC be a triangle and let the points K,L,M be on the lines that contain the sides BC, CA and AB respectively. (Note that at least one of the sides must be produced.) Then K,L,M are collinear if and only if

$$\frac{\overrightarrow{AM}}{\overrightarrow{MB}} \cdot \frac{\overrightarrow{BK}}{\overrightarrow{KC}} \cdot \frac{\overrightarrow{CL}}{\overrightarrow{LA}} = -1.$$



Proof: Here we only indicate the proof. For the first part assume that K, L, M are on a line and drop perpendiculars from the vertices A, B, C to this line. By a similar procedure to that of the proof of Ceva's theorem, find similar triangles to establish the identity first without considering the signs of the ratios. Next note that while $\frac{\overrightarrow{AM}}{\overrightarrow{MB}}$ and $\frac{\overrightarrow{CL}}{\overrightarrow{LA}}$ are positive, $\frac{\overrightarrow{BK}}{\overrightarrow{KC}}$ is negative, since BK is oppositely directed to KC. (Such a situation occurs always. Either one or three intersections must be on the sides produced.) The converse part of the proof you obtain by a reasoning similar to that for Ceva's theorem, that is, by assuming the identity and considering the point K' where ML produced intersects BC produced.

We add a short note on the history of these important and useful theorems. Menelaus was an astronomer who lived in the Alexandrian period of the Greek civilization (about 100 A.D.). Ceva's contribution to geometry came much later (about 1680).

* * * * *

HOW MUCH MATHEMATICS CAN THERE BE?

With billions of bits of information being processed every second by machine, and with 200,000 mathematical theorems of the traditional, hand-crafted variety produced annually, it is clear that the world is in a Golden Age of mathematical production. Whether it is also a golden age for new mathematical ideas is another question altogether.

It would appear from the record that mankind can go on and on generating mathematics. But this may be a naive assessment based on linear (or exponential) extrapolation, an assessment that fails to take into account diminution due to irrelevance or obsolescence. Nor does it take into account the possibility of internal saturation. And it certainly postulates continuing support from the community at large.

It seems certain that there is a limit to the amount of living mathematics that humanity can sustain at any time. As new mathematical specialties arise, old ones will have to be neglected.

All experience so far seems to show that there are two inexhaustible sources of new mathematical questions. One source is the development of science and technology, which make ever new demands on mathematics for assistance. The other source is mathematics itself. As it becomes more elaborate and complex, each new, completed result becomes the potential starting point for several new investigations. Each pair of seemingly unrelated mathematical specialties pose an implicit challenge: to find a fruitful connection between them.

The Mathematical Experience,
Philip J. Davis and Reuben Hersh, Penguin, 1980.

CENTRIFUGAL FORCE

Michael A.B. Deakin and G.J. Troup
Monash University

1. *History and General Principles*

The question of *centrifugal force* is one that often gives rise to controversy and needless confusion. Often, Physics texts ignore it, or seem to replace it with a different force called a *centripetal force.* Now "centrifugal" means "centre-fleeing", while "centripetal" means "centre-seeking", so the two words refer to precisely opposite directions. Students are thus easily misled into thinking that these forces are antagonistic to each other, or that centrifugal force is a fiction (despite the fact that we all experience it).

The true picture is quite simple, and we present it here.

Let us begin with a very common early childhood experience. Most of us, perhaps all of us, when young, saw street-trees move backwards as we sat in a moving vehicle. And very likely we were told: "No, dear, the trees aren't moving backwards, *we're* moving forwards".

Well, as Einstein would have said, it all depends on your point of view.

Newton, in his analysis of motion, saw the importance of the viewpoint one takes when one observes the motion. We now speak of the *frame of reference* adopted by the observer. Now Newton held that the universe itself provided us with a uniquely preferred frame of reference – one that was better than all others in the analysis of mechanical phenomena. For he believed that an *absolute space* was embedded in the structure of the universe.

We do not see absolute space, but Newton, and physicists in general for many years, saw it as revealing itself through its effects. If we look at the night sky, we see that the stars move, but in a regular way, keeping their patterns of constellations fixed. The sun, moon and planets (and the earth too) move with respect to these "fixed stars". Newton saw the fixed stars as giving us the reference frame supplied by his absolute space.

For Newton, therefore, absolute motion was motion relative to absolute space. As he wrote, "Absolute motion is the translation of a body from one absolute place to another". It is these absolute motions that obey Newton's laws.

Nowadays we are less rigid about these matters. Since Einstein, we recognise that all frames of reference have equal validity, although some, for one reason or another, are more convenient to use. Among these frames there are some – the so-called *inertial frames* – that are particularly useful. The frame defined by the fixed stars is inertial, as is any frame in uniform motion (i.e. moving in a straight line at constant speed) with respect to this. These are the only inertial frames that can exist.

Newton's laws hold (to a very good approximation) in inertial frames, but not in other, non-inertial ones.

We usually think of a body being at rest if it is at rest relative to the earth. So we implicitly use the earth as our frame of reference. This is not an inertial frame, although for many practical purposes we may think of it as being so.

However, if our frame of reference is the earth, its non-inertial character will lead to departures from Newton's laws. The most obvious of these is the motion of the sun across the sky. A purely terrestrial example is the motion of the Foucault pendulum, such as that in Monash's mathematics building. (See *Function, Vol.* 6, *Part* 2.)

We tend to say: "The sun's movement across the sky is an apparent motion caused by the rotation of the earth". But we are too glibly accepting the idea of an absolute space if we do this. Modern physics, since Einstein, sees no frame as being superior to any other.

Even before Einstein, the mathematics of non-inertial frames was investigated. We shall make particular reference to those frames which, like the earth, are rotating. These developments were due to Huyghens (1629-1695), a Dutch physicist and mathematician and, in a fuller and more developed form, to Coriolis (1792-1843) and Foucault (1819-1868) who were French.

2. *Analysis*

Consider the case of a (top-loading) spin-dryer, as shown in Figure 1. A water droplet inside an article of clothing is initially at P, but it leaves the dryer at that time. We will analyse from two points of view what happens.

First, consider the matter from the Newtonian perspective. The spin-dryer is pivoted at O, which is at rest relative to the earth (which we will treat as an inertial frame, as the minute effects of its motion are unimportant in this case). The drum of the dryer rotates about O with an angular velocity ω radians per second, so that in time t, P moves to the new position P' and $\angle POP' = \omega t$. *The water droplet*, meanwhile, is acted on by no forces (if we neglect gravity and air-resistance) and so obeys Newton's first law. It thus proceeds to travel in a straight line with constant velocity.

Now when it left the dryer, at P, the drop was travelling at the same velocity as P. From Figure 1, we see that P had a velocity $a\omega$ (a being the radius) in a direction perpendicular to OP. The drop continues to travel with this velocity and so reaches a point Q, distant $a\omega t$ from P, after time t has elapsed.

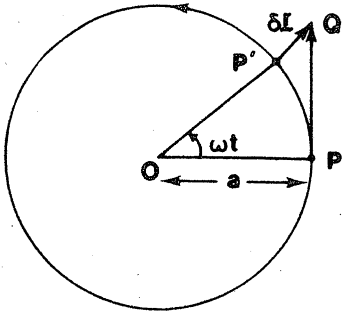


Figure 1

Introducing a unit vector $\underset{\sim}{i}$ in the direction $\overrightarrow{OP}$ and a unit vector $\underset{\sim}{j}$ in the direction $\overrightarrow{PQ}$, we now see that the drop has position vector, relative to O, $\underset{\sim}{i}a + \underset{\sim}{j}a\omega t$, while the piece of clothing it has left, now has the position vector $\underset{\sim}{i}a \cos \omega t + \underset{\sim}{j}a \sin \omega t$.

On this account, we have no need to invoke centrifugal force. What requires explanation is the circular path adopted by the clothing, and this is due to a centri*petal* ("centre-seeking") force in the direction $\overrightarrow{P'O}$. This is supplied by reaction with the wall of the dryer, and it explains why the clothing does not do the natural thing and travel with constant velocity, i.e. in a straight line with constant speed.

But now look at the matter differently. Imagine an observer travelling with the item of clothing. *This* observer sees the drop leave the wall of the dryer and travel *out* to Q. If we write $\underset{\sim}{r}$ for the vector $\overrightarrow{OP'}$ and $\delta\underset{\sim}{r}$ for the vector $\overrightarrow{P'O}$, we have

$$\delta\underset{\sim}{r} = \underset{\sim}{i}a + \underset{\sim}{j}a\omega t - \underset{\sim}{i}a \cos \omega t - \underset{\sim}{j}a \sin \omega t.^{\dagger}$$

---

† This equation can be recast to give two equations in coordinates X (out along $\overrightarrow{OP}$) and Y (perpendicular to this). We then have (try it as an exercise)

$$\left.\begin{aligned} X &= a(\cos \omega t + \omega t \sin \omega t - 1) \\ Y &= a(\omega t \cos \omega t - \sin \omega t) \end{aligned}\right\}.$$

These equations give the curve known as the *involute of a circle*, produced by the end of a piece of thread held taut as it is unwrapped from a spool. Can you see why this should be so?

For this observer, Q will have an acceleration of $d^2(\delta \underset{\sim}{r})/dt^2$, i.e. $\omega^2(\underset{\sim}{i}a \cos \omega t + \underset{\sim}{j}a \sin \omega t)$ or $\omega^2\underset{\sim}{r}$. Notice that this is an outward acceleration and so is termed a centri*fugal* (or "centre-fleeing") acceleration.

In this view of things, our observer sees no need to explain his own motion, since he is at rest relative to himself. But he does need to explain the acceleration of the particle. Now of course he *could* say: "I'm not really in an inertial frame and this acceleration is only apparent and is caused by my motion". But equally well he could say this: "The acceleration must be caused by a force $\underset{\sim}{F}$ and, by Newton's second law, this is $m\omega^2\underset{\sim}{r}$, where m is the mass of the drop". Such a force is termed *centrifugal force*.

3. *Rotating Frames in General*

This is only one, very simple, example of a rotating frame. It was once possible to experience something very like this at Luna Park in Melbourne. A large machine called a Rotor (and very much resembling a spin-dryer) was located there, and people could enter it and be spun around so that they were pinned to the walls. At this point, the floor was dropped so that the adventurers were as shown in Figure 2, in equilibrium (more or less) under the influence of four forces: N, the normal reaction of the wall; C, the centrifugal force; W, their weight; F, the friction.

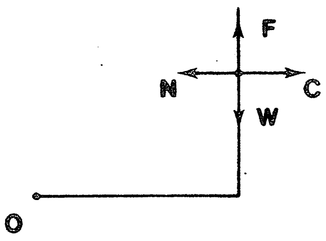


Figure 2

Some hardy souls used to go into the Rotor in groups and try to throw tennis balls to one another. Observers above, who were not themselves rotating, would then see the Newtonian view, while those inside the contraption saw matters from a non-inertial perspective. The real skill was to throw the ball to yourself.

Figure 3 gives two views of this. A person at P, in Figure 3(b), throws the ball, not straight out, say, but rather against the rotation. However, this does not entirely overcome the effect of the rotation, so that the stationary observers see the motion $\overrightarrow{PP'}$ of Figure 3(a), as followed by the ball, while the thrower travels from P to P′ in a curved path and so catches the ball.

The thrower, however, sees the ball travel out and loop back, as shown in Figure 3(b). The ball is forced back to his hand.

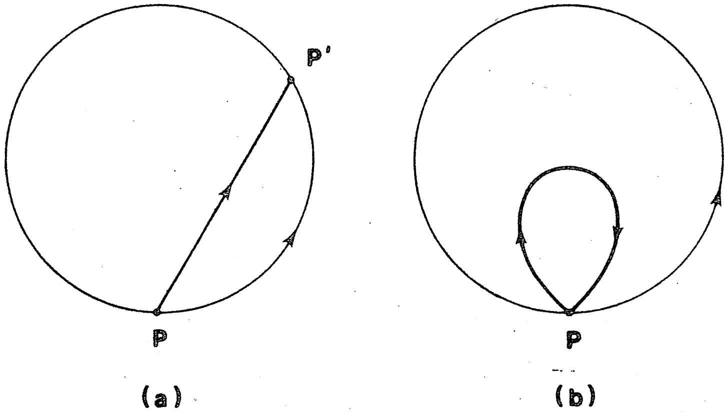


Figure 3

This is a much more complicated situation than the earlier one. There are in fact two forces introduced by the fact that the frame is rotating. These are: the centrifugal force, which we have already seen, and the so-called *Coriolis force*, which is due to the fact that the ball is moving with respect to the rotating frame.

Once it leaves the spin-dryer, the water-drop is also moving with respect to the rotating frame and so is subject to a Coriolis force. This explains why $OP'Q$ is not straight (as one might perhaps expect) in Figure 1.

Because the earth is often used as a frame of reference and this is rotating, Coriolis forces become important in meteorology and oceanography. (See *Function, Vol.* 1, *Part* 2; *Vol.* 1, *Part* 4; *Vol.* 2, *Part* 3.) Ballistics is another area where they come in, and so is the theory of the Foucault pendulum. Centrifugal force enters into the theory of planetary motion (see *Function, Vol.* 8, *Part* 5), and Newton himself used the notion in his account of the shape of the earth in the third volume of his *Principia*.

4. *Modern Developments*

There are two quite distinct notions of the mass of a body:

(1) mass may be thought of as *inertial mass*, a measure of the body's accelerative response to an impressed force;

(2) mass may be thought of as *gravitational mass*, a measure of the body's ability to set up a gravitational field.

However, while this distinction is quite clear and may be traced back to Newton himself, experiment has shown that to great accuracy, the two values turn out to be equal.

This led Einstein to develop the *Principle of Equivalence* (see *Function, Vol.* 3, *Part* 4) according to which gravitational forces may be explained as due to the non-inertial nature of the frame in which they are observed.

To see this, consider a rider on a merry-go-round. See Figure 4. In the rotating frame, the rider is in equilibrium as the tension T in the support exactly balances the weight W and the centrifugal force C.

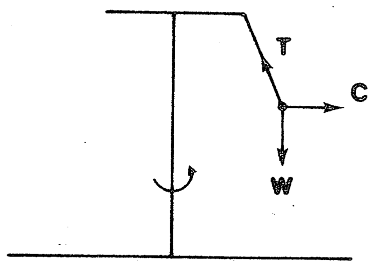


Figure 4

Some authors see these as forming *two components* of the rider's weight, the centrifugal force being a horizontal component of gravity. This kind of language is used (for example) in connection with the space programme. Consider the frame of reference attached to an orbiting space-craft. The earth-directed force of gravity is balanced by the centrifugal force, so that the astronaut is "weightless".

What Einstein did was to suggest that what we normally term "gravity" was also, in effect, caused by the frame of reference, just as the centrifugal force is.

Centrifugal and Coriolis forces are determined by the inertial mass, rather than the gravitational mass. They thus tend to be called *inertial forces* (a much better name than that adopted by some authors, who call them "fictitious forces" – they are quite real, we all experience them). Einstein thus suggested that gravity itself was really also an inertial force.

Consider the surface of a liquid (Figure 5) held in a vessel, such as a bucket. If the bucket is at rest (in an inertial frame), its surface will be flat, at right angles to the force of gravity W. This is shown in Figure 5(a). If the liquid is rotating, its surface acquires a parabolic shape as shown in Figure 5(b), as the weight now has a horizontal component due to the centrifugal force. The surface is still everywhere perpendicular to the weight force W acting on the particles that compose it.

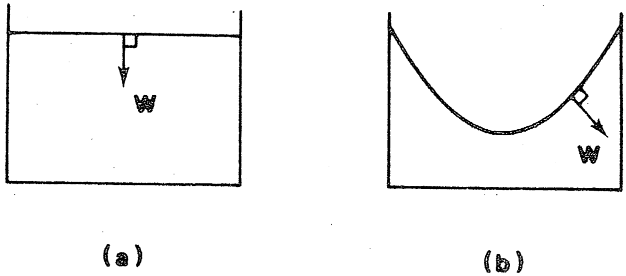


Figure 5

This example is due to Newton, who used the shape of the surface to distinguish inertial from rotating frames (although he didn't put it quite that way). Einstein reversed this view. The surface is paraboloidal because it is still, if not flat, then as flat as is possible, in a curved non-Euclidean space. Similarly, planets travel in elliptical (or nearly elliptical) orbits, because the sun warps the space in which they travel and then inertial forces make them go in the straightest possible lines in this space. (See *Function, Vol. 3, Part* 2.) They are not regarded as being subject to gravitational forces, but rather as merely obeying an extended version of Newton's first law.

If the notion of curved space seems strange, recall that we ourselves live on the curved surface of the earth and may not proceed from one place to another by straight paths. The straightest possible paths are so-called "great circles" and they are much used in navigation. (See *Function, Vol. 4, Part* 1; *Vol. 6, Part* 4; *Vol. 6, Part* 5.) Airliners, for example, follow them.

Einstein suggested that the planets, especially Mercury, acted like airliners and chose these straightest paths, in a space warped by the presence of the sun. As is now known, this led to an explanation of observed discrepancies between the actual and the theoretical positions of this planet.

One problem Einstein was unable to come to grips with completely. It still remains, and from time to time different researchers claim success, but no one theory has achieved universal acceptance. This is to explain why the frame of reference defined by the fixed stars should be inertial. It seems as if inertia – the tendency of a body to obey Newton's first law – is somehow a property brought about by the matter great distances away, or, if you like, by the distribution of matter in the universe as a whole. But it is not known how this connection works.

* * * * *

INDEX, VOLUME 12