Mathematical Spectrum

Volume 10 1977/78 *Number 2*

A Magazine of *Contemporary Mathematics*

Published by the *Applied Probability Trust*

Mathematical Spectrum is a magazine for the instruction and entertainment of student mathematicians in schools, colleges and universities, as well as the general reader interested in mathematics. It is published by the Applied Probability Trust, a non-profit making organisation established in 1963 with the support of the London Mathematical Society. The object of the Trust is the encouragement of study and research in the mathematical sciences.

Volume 10 of *Mathematical Spectrum* will consist of three issues of which this is the second. The first issue was published in September 1977, and the third will appear in May 1978.

Articles published in *Mathematical Spectrum* deal with the entire range of mathematical disciplines (pure mathematics, applied mathematics, statistics, operational research, computing science, numerical analysis, biomathematics). Both expository and historical material may be included, as well as elementary research and information on educational opportunities and careers in mathematics. There is also a section devoted to problems.

Articles are normally commissioned by the Editors; the Editorial Committee also welcomes the submission of suitable material, including correspondence, queries and solutions to problems, for publication in *Mathematical Spectrum*. All correspondence about the contents should be sent to:

The Editor, *Mathematical Spectrum*,
Hicks Building, The University, Sheffield S3 7RH.

Mathematical Spectrum Awards for Volume 9

The editors have awarded the prize of £20 to Patrick Brooke for his article 'Perfect and Pandiagonal Magic Hyper-Cubes' (Volume 9, Number 3, pp. 82–94); and the prize of £10 to Alan Burns for his solution of Problem 9.3 (Volume 9, Number 3, pp. 98–99).

The 19th International Mathematical Olympiad

T. J. HEARD
City of London School

At Belgrade, in July 1977, 156 pre-university students from 21 countries competed in the 19th International Mathematical Olympiad. Each delegation submits problems in advance, which are considered by all the leaders just before the competition, when a final selection of six is agreed. On two successive days the competitors tackle three problems in four hours. Their scripts are assessed first by the team's leader and his deputy, who then discuss the solutions of each problem with two Yugoslav coordinators whose task is to maintain a common standard for that one problem. The few cases where coordinators and leaders disagree on the value of a solution are considered by the whole jury, a wearying process when conducted in five languages on a hot afternoon.

The total scores (maximum 320) and prizes for each team of eight were as shown in Table 1.

This year's British team had only one overlap with last year's, and included two representatives from Scotland and one from Northern Ireland. The team's position has been bettered only once in the eleven years in which we have taken part (second in 1976, though then we were 36 points behind the winners, compared with 12 points this year). One individual deserves particular mention: John Rickard (City of London School), a team member for the third time, scored full marks and received two special prizes for generalisations of problems. This unique achievement was rightly acclaimed at the closing ceremony.

Of course there is much more to an international gathering like this than the hard grind of the competition. Although the teams and leaders came together only for the last two days, when assessment had been completed, both groups were well entertained by our generous and efficient Yugoslav hosts, with excursions, receptions, concerts, a simultaneous chess display (where Richard Borcherds of King Edward's School, Birmingham, was one of only three to beat the Yugoslav grand master), and time for making friends and learning about other countries.

The degree of national support which the teams receive varies widely. The selection of teams always seems to start with regional or national competitions, such as our National Mathematics Contest and British Mathematical Olympiad, though often these competitions are much more part of the normal curriculum than they are here. We were told that the selection of this year's teams had involved some three million pupils. Some countries (e.g. USSR and Yugoslavia) have special mathematical schools catering for those who are successful. Others select a pool of 'possibles' who are given training either by correspondence or at regular weekend or holiday sessions, and from whom the team is chosen. Of all the full teams the British, who met for the first time at Heathrow, seem to have the least preparation. This makes their good results all the more praiseworthy, but it also prompts one to wonder whether with a modest extra effort we could not win.

TABLE 1

Place		Total	Prizes 1st	2nd	3rd	Special
1	USA	202	2	3	1	
2	USSR	192	1	2	4	1
3=	Great Britain	190	1	3	3	2
3=	Hungary	190	1	3	2	1
5	Netherlands	185	1	2	3	
6	Bulgaria	172		3	3	
7	West Germany	165	1	1	4	
8	East Germany	163	2	1	1	1
9	Czechoslovakia	161		3	2	1
10	Yugoslavia	159		3	3	
11	Poland	157	1	2	2	
12	Austria	151	1	1	2	
13	Sweden	137	1	1	2	
14	France	127	1			
15	Rumania	122		1	2	
16	Finland	88			1	
17	Mongolia	49				

In addition, teams of fewer than eight from Cuba, Belgium, Italy and Algeria came 18th to 21st respectively. Greece and Vietnam had to withdraw at the last moment.

Appendix

These were the problems for 1977.

1. (6 points) Equilateral triangles ABK, BCL, CDM, DAN are constructed inside the square $ABCD$. Prove that the midpoints of the four segments KL, LM, MN, NK and the midpoints of the eight segments AK, BK, BL, CL, CM, DM, DN, AN are the twelve vertices of a regular dodecagon.

2. (6 points) In a finite sequence of real numbers the sum of any seven successive terms is negative and the sum of any eleven successive terms is positive. Determine the maximum number of terms in the sequence.

3. (7 points) Let n be a given integer > 2, and let V_n be the set of integers $1 + kn$, where $k = 1, 2, \ldots$. A number $m \in V_n$ is called *indecomposable in* V_n if there do not

exist numbers $p, q \in V_n$ such that $pq = m$. Prove that there exists a number $r \in V_n$ that can be expressed as the product of elements indecomposable in V_n in more than one way. (Expressions which differ only in the order of the elements of V_n will be considered the same.)

4. (6 points) a, b, A, B are given constant real numbers, and

$$f(\theta) = 1 - a\cos\theta - b\sin\theta - A\cos 2\theta - B\sin 2\theta.$$

Prove that, if $f(\theta) \geqslant 0$ for all real θ, then

$$a^2 + b^2 \leqslant 2 \qquad \text{and} \qquad A^2 + B^2 \leqslant 1.$$

5. (7 points) Let a and b be positive integers. When $a^2 + b^2$ is divided by $a + b$ the quotient is q and the remainder is r. Find all pairs (a, b), given that $q^2 + r = 1977$.

6. (8 points) Let $f(n)$ be a function defined on the set of all positive integers and taking on all its values in the same set. Prove that, if $f(n + 1) > f(f(n))$ for each positive integer n, then $f(n) = n$ for each n.

Isaac Newton—an Anniversary

HAZEL PERFECT
University of Sheffield

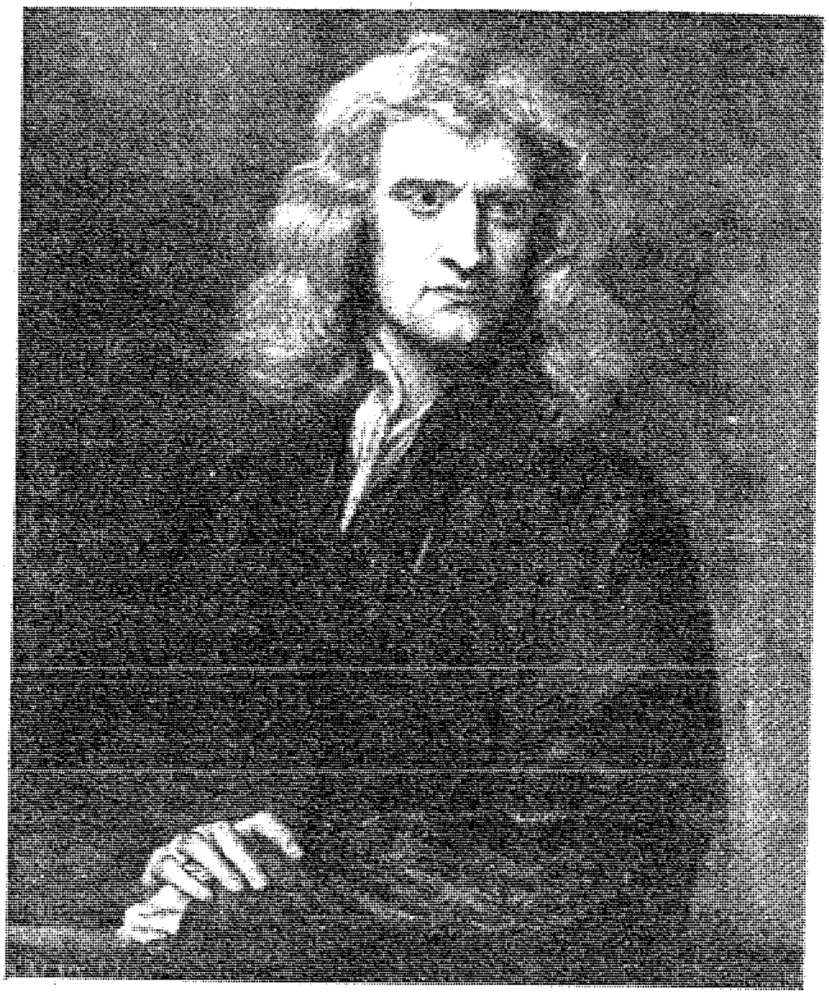
Isaac Newton died just over two hundred and fifty years ago on 20 March 1727. Let us not allow *Mathematical Spectrum* to leave unrecorded the anniversary of this great man, supreme among mathematicians, who ascribed his achievements to the fact that he had 'stood on the shoulders of giants'.

Newton indeed had giants among the men of science preceding him: in the forefront, Galileo, Kepler, and Descartes. Galileo Galilei (1564–1642), a native of Pisa in Italy and later to hold academic posts in the Universities of Pisa and Padua, may justly be said to have inaugurated modern physical science. We recall that, while still a student, Galileo discovered the laws governing the oscillations of a pendulum. He also constructed the first telescope and discovered the four outer satellites of Jupiter. Galileo's studies led him to adopt the Copernican view of the universe, in which the planets revolve around the fixed sun as centre; but, when opposition within the Church to the new cosmology hardened and he was summoned to Rome in 1633, he was prevailed upon to sign a retraction of his views. It is of particular relevance for us to record here Galileo's hypothesis that: *a body upon which no external force is at work maintains its original velocity* (*or state of rest*). In the second of his three basic laws of motion, Newton was later able to extend and quantify this statement and to formulate with precision the notions of acceleration and force. In a short essay written in 1949 on the German astronomer Johann Kepler (1571–1630), the great Einstein paints a picture of a man of 'sensitive personality, passionately devoted to the quest for deeper insight into the character of natural processes, [who had to suffer] poverty and lack of comprehension among those of his contemporaries who had the power to shape his life and work'. Nevertheless, to Kepler the world owes the empirical discovery of the laws describing the motions of the planets: *the planets move in ellipses with the sun at a focus; equal areas are swept out in equal times by a line joining the planet to the sun; and the square of the time of a complete revolution is proportional to the cube of the planet's average distance from the sun.* It was Newton who later provided their mathematical derivation. Kepler was of an imaginative disposition; in contrast to the obsessively reticent Newton, he was very vocal and ecstatic about his discoveries, and they evidently gave him intense pleasure and excitement, which he made no attempt to hide. René Descartes (1596–1650), French philosopher and mathematician, came of an old noble family and, being a delicate child, had little formal early education, and later met with much special consideration during his schooldays at a Jesuit school. Thus he was allowed ample time to follow his own inclinations, and we read of long quiet mornings in bed spent in the study of mathematics and philosophy. Later in life, rather paradoxically, he enlisted in the army to escape the distractions of the social round and find the tranquillity he needed for his work. Still in an age of universalism in science, Descartes studied not only mathematics and philosophy but also most of the sciences of his day. He made decisive advances in philosophy, but his greatest contributions were in pure mathematics; and his invention of coordinate geometry probably provided Newton with an effective mathematical tool.

We now turn to Newton himself. Born on Christmas Day 1642, Isaac Newton was the son of a farmer who owned the Manor of the small village of Woolsthorpe near Grantham in Lincolnshire. He was a posthumous child and so small and sickly at birth that he was not expected to survive. He did, however, live into his eighty-fifth year. During his formative years of childhood, following his mother's remarriage, Newton was brought up by his grandmother under his uncle's guardianship, and from the age of two years he had settled upon him the quite substantial annual income of £80. The text books of William Oughtred, who first introduced the familiar × sign of multiplication and who invented the slide rule, were those used by the young Isaac Newton and from which he acquired his early grounding in mathematics. It was intended that the boy should follow his father's vocation and become a farmer. He did not, however, show any inclination for the work and spent his spare time devising mechanical toys and performing scientific experiments as well as drawing and painting. Newton entered Grantham Grammar School and there received a thoroughgoing classical education. His scientific interests were satisfied, however, by the chemistry books which he found in the attic of his apothecary friend with whom he lodged in Grantham. From the Grammar School, Newton proceeded, at the age of eighteen, to Trinity College, Cambridge. His teacher there was Dr Isaac Barrow, the first holder of the Lucasian Chair of Mathematics. During his first two years as an undergraduate, Newton made no particular impression. But Barrow was one of the foremost mathematicians of his period and, moreover, his own researches were leading him in the direction of the calculus, so that his influence upon Newton began to make itself felt and eventually set Newton's own genius on fire. Kepler's 'Optics', the mathematical works of Euclid, Apollonius, John Wallis and others, the analytical geometry of Descartes were all absorbed by Newton. Thus his technical proficiency was already more than adequate, and his mind was seething with ideas. It was always Newton's predilection to work on his own and to seek to gain insight by himself; for him, mathematical science was a private activity. After three years at Trinity, Newton graduated from there in 1665 and was awarded a Minor Fellowship to enable him to pursue his studies.

We have already hinted at Newton's attitude to his work. He worked alone and cared little to share his knowledge. His joy was in creating, and there was, at the same time, a minute precision in all that he did. Later in his career, this feeling for science as something personal and private was reflected in his reluctance to publish his work. It appears that he felt no obligation to Trinity College to communicate his great discoveries by a sense of gratitude for the opportunities that had been provided for him. In 1665, the bubonic plague forced Newton to leave Cambridge for a period of some eighteen months; but, for a man of his temperament, this provided the perfect conditions for private study in his favourite surroundings in Lincolnshire, with no distractions to come between him and the objects of his thoughts. There he used prisms to investigate the fundamental properties of light; and having already absorbed the Copernican theory of the universe, he also pursued further his work in astronomy. (Later, he was to construct the first reflecting telescope.) It is from this period that we have the famous—and probably true—story of the apple; and undoubtedly the theory of gravitation was shaping in his mind and had been virtually formulated before his return to Cambridge.

Newton was elected to a Major Fellowship of Trinity College in 1667 and he gave his first lectures there in optics. In 1669, Barrow took the unusual step of relinquishing his mathematical chair in favour of his more gifted student. From this appointment, Newton received a further annual income of £100 and relatively few formal duties: he was to give twenty-four lectures annually and to supervise a small number of students. At this period, or a little earlier, we find Newton at the height of his powers—indeed with practically all his greatest mathematical and scientific work accomplished in essence or, at least, foreshadowed. Let us pause for a moment to consider something of what Newton had, in fact, achieved. His three laws of motion had been formulated: 1. originally stated by Galileo and italicised above; 2. *the acceleration of a body is proportional to, and in the same direction as, the applied force;* 3. *to every action there is an equal and opposite reaction.* What was especially significant was Newton's explicit recognition of the universal applicability of these laws, i.e. that they governed the motions of the heavenly bodies as much as the motions on this planet: truly a revolution in scientific thought. The differential and integral calculus had been forged as tools for a searching investigation of such notions as velocity, acceleration, area and so on. Kepler's laws, in particular that concerning the elliptical orbits of the planets, had been mathematically justified as a consequence of the second law of motion and the general theory of gravitation



Portrait of Newton by Kneller, 1689.

based on the inverse-square law. (Further spectacular evidence for the truth of the theory of gravitation was later to be provided by observations on Halley's comet.) By applying general principles to motions in the heavens, the master intellect had created the subject of celestial mechanics. In the field of optics, Newton had made the discovery that white light can be separated into individually recognizable rays corresponding to the colours of the spectrum.* All this varied and profoundly original work had been done before Newton had reached the age of twenty-five. Official recognition soon followed, and in 1672 he became a fellow of the Royal Society. It was not until 1684 that Newton, at the urgent request of his friend Edmond Halley, began to write up his scientific work in his now famous 'Philosophiae naturalis principia mathematica' (Mathematical Principles of Natural Philosophy†). In less than two years the 'Principia' was completed and presented to the Royal Society; and it was published in 1687. The record of

* Whereof the cover of this magazine is a visible reminder.
† Today, we speak of 'physics' rather than 'natural philosophy'.

Newton's scientific life is marred by his controversies with Hooke over their researches into the nature of light and colour, and with Leibniz over the discovery of the calculus. As a result of the first, Newton refused to accept the Presidency of the Royal Society until the death of Hooke in 1703. The second had its source in Newton's aversion to publication. The calculus had been in Newton's grasp since before 1665 but not in print, whereas Leibniz had already published his own account. This had the added advantage of a good notation, which stimulated further research on the subject on the Continent. Newton's mathematical acuity lasted until late in life although he gave up his academic career at the age of fifty-four upon his appointment as Warden (later Master) of the Mint.

The mathematical story has now been told; the rest is subordinate. Throughout his life, Newton maintained an interest in chemistry and theology. The suggestion has been advanced that these were his main preoccupations, and it appears that he spent more time on them than on his mathematical physics. His intense labours in chemistry, however, produced relatively few results. Newton left Cambridge and scientific research with no apparent regrets when he went to the Mint, and he only rarely revisited Cambridge. Newton never married. Until the end of his life, he exercised his duties as Lord of the Manor of Woolsthorpe with meticulous care; he remained a life-long member of the Church of England although, on theological grounds, he declined to be ordained, and thus made himself ineligible for election as Master of Trinity College. In 1705, Newton was knighted by Queen Anne. Though he died a rich man, and in spite of earlier intentions, Newton left no financial endowments to any scientific institution.

A brief enough outline of the life of a genius,* but a more detailed catalogue of Newton's achievements or an attempted assessment of his work within the compass of a few more pages, could still scarcely begin to measure its importance. Newton, too, had broad shoulders, and giants of a later age have stood upon them to produce some of the spectacular advances in mathematics and physics since his death. We, who live in the post-Newtonian age, today salute his memory.

* There is a wealth of material available about Isaac Newton. We refer the interested reader to Chapter 6 in E. T. Bell's *Men of Mathematics* (Dover Publications, 1937), to sections in the same author's *Development of Mathematics* (McGraw Hill, 1945); and especially to the biography *Sir Isaac Newton* by E. N. da C. Andrade (in the series 'Brief Lives', Collins, 1954) written with beautiful simplicity and clarity.

How Many Bird Territories are there on a Farm? A Statistical Approach to an Ornithological Problem

PHILIP M. NORTH
University of Kent, Canterbury

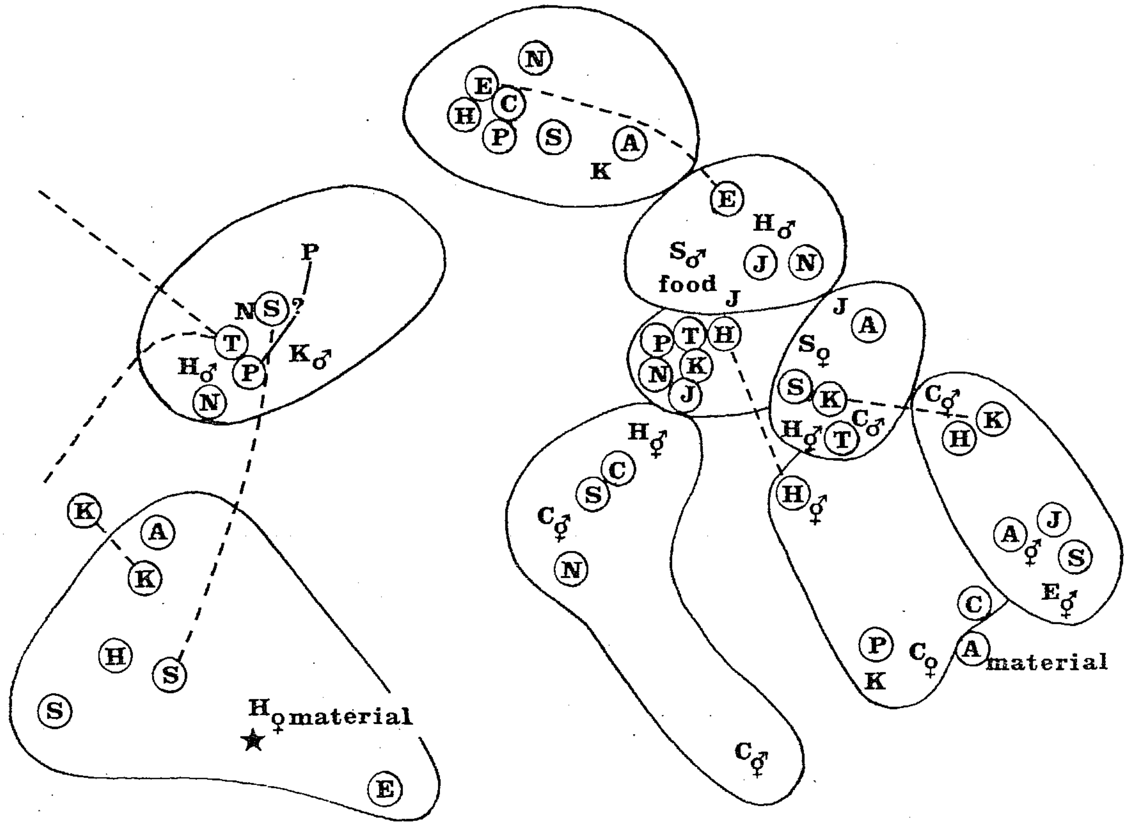
1. Introduction

Since 1962 the British Trust for Ornithology (B.T.O.) has conducted an annual Common Birds Census to monitor fluctuations in the population levels of Britain's commoner birds. It was originally instigated by the Nature Conservancy because of concern over the possible harmful effects on wildlife of modern farming methods, including the use of insecticides and herbicides. The census was therefore initially carried out on farmland, but its scope has since been extended to cover other types of habitat as well, notably woodland. So far, farmland data only have been used in the cluster analysis described below.

2. The fundamental stages of the Common Birds Census

The fieldwork for the census is undertaken by a large number of voluntary observers, who form part of the vast body of amateur ornithologists which now exists in Britain. The analysis of the data collected leads to the calculation of index values for each species considered, which indicate species population sizes relative to 1966 as base year. This year was chosen as it was considered that most species had by then made up the losses which many had incurred during the severe winter of 1962–63. The basic units required for the calculation of these index values are the numbers of territories held each year. But bird territories cannot easily be counted in the field, and for most species the number present on any plot of land can in practice only be estimated. This is done using an internationally approved mapping method for census work. For this, each observer is required to visit one plot of land (for farmland, typically about 75 hectares in area) regularly throughout the breeding season; ideally, each should make at least nine visits, and certainly at least six. On each visit the observer registers in the correct positions on a map of the plot the birds encountered on that visit, together with descriptive information, such as 'male bird singing'. These maps are then used to compile separate species maps displaying the information gathered from all the visits; the order of these is noted on the maps by using successive letters of the alphabet for the registrations (see Figure 1). Approximate territories are then located by B.T.O. staff by looking on the maps for distinct groups of registrations from different visits, also taking into account the descriptive information and the various habitat types within the plot, e.g. hedges, open fields.

This method of analysing the data has two main disadvantages. First, it is vastly time-consuming, as B.T.O. staff members have to analyse separately, by hand, every species map from every plot; there are about 3600 maps for farmland alone each year. Second, the assignment of territories may be partly subjective, in spite of the carefully defined procedure for analysis.



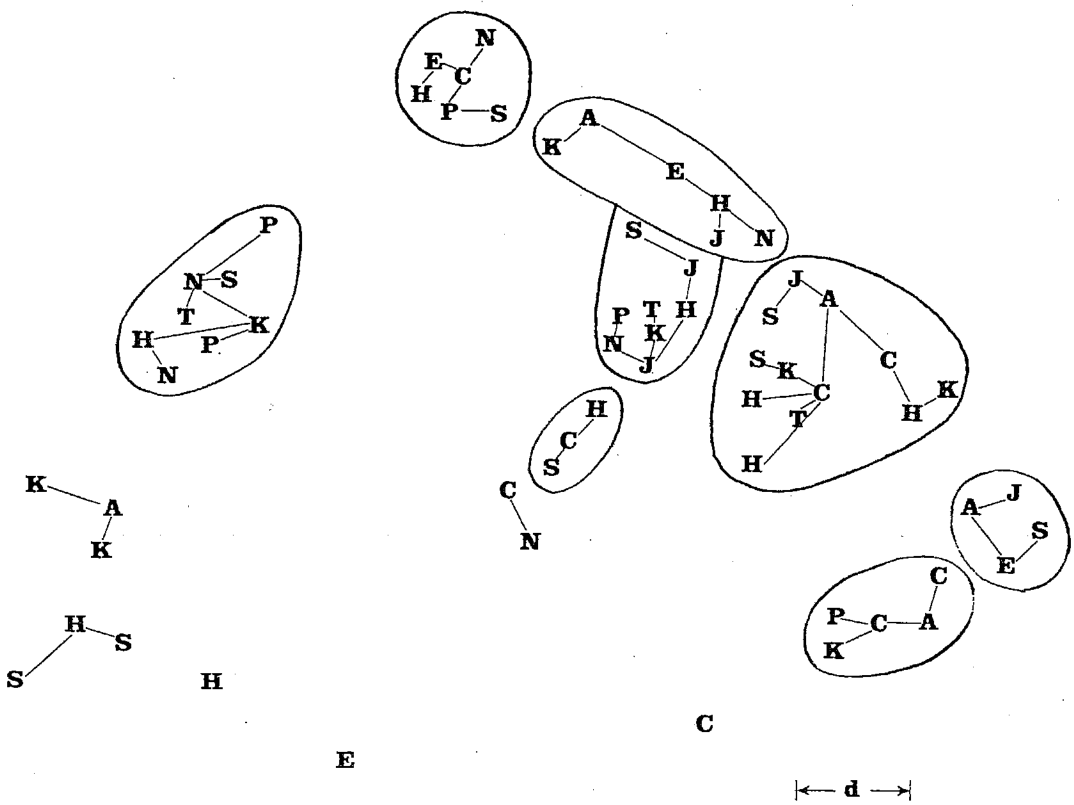
KEY:		
	A	Singing male
	A material	Seen with nest material in beak
	A food	Seen with food in beak
	★	Nest
	A♂ , A♀	Sight-record, with indication of sex if appropriate
	A - - - - - - - - A	Different birds heard in song at the same time
	A ———— A	Singing bird seen to take up a new position
	A —— ? —— A	Thought to be the same bird in a changed position but not certain

The different letters in the figure denote the different visits, in alphabetic order, which produced the registrations, *A* corresponding to the first visit, *B* to the second, and so on. The enclosing lines denote approximate territory boundaries, as marked by B.T.O. staff.

Figure 1. Chaffinch registrations with B.T.O. territory assessment.

3. An alternative approach: a new method of cluster analysis

There is clearly a need for a quicker, easy to use, fully objective method, either as a check on the present procedure or as an alternative method. It is quite possible that to obtain estimates of just the numbers of territories on plots rather than their actual demarcations, the B.T.O. method uses more information than is necessary. As an initial step, the data have therefore been drastically simplified by ignoring the descriptive information, to produce the data illustrated in Figure 2, showing just the positions and timing of registrations. Interest lies in the amount of progress that statistical techniques enable us to make with this revised set of data.



Lines joining the registrations represent links made in the cluster analysis performed using the defining distance shown. Enclosing lines are used simply to distinguish between clusters satisfying the minimum requirements for contribution to the estimated number of territories, and do not necessarily represent territory boundaries.

Figure 2. Simplified Chaffinch registrations with a clustering territory assessment. The defining distance d is given at the bottom right-hand corner.

The mathematical problem is now one of cluster analysis, that is grouping together registrations of birds of the same species to define their territories, using standard multivariate statistical methods. The particular interest of this problem lies in the heterogeneity of the points to be clustered, for we need to group together, in a sensible way, points which are associated with different instants of time. There already exist statistical techniques for clustering a set of homogeneous points and the method which has been developed for the present problem is essentially an extension of one of these, called single linkage cluster analysis. The reader interested in a review of methods of cluster analysis, and further references, should consult the book by Everitt (reference 1).

4. Rules for the new method

The clustering of the bird registrations is carried out according to the following set of rules:

(i) Specify a defining distance d for a cluster. Since the cluster method has so far been used only as a check against the B.T.O. analysis, the clustering has been performed in each case for a sequence of defining distances d_1, d_2, d_3, . . . , and the results compared with the B.T.O. one;

(ii) Link each B registration to its nearest A registration if they are no further than distance d apart;

(iii) Link each C registration to its nearest A or B registration (the closer) if they are no further than distance d apart;

(iv) Link each D registration to its nearest A, B or C registration (the closest) if they are no further than distance d apart;

and so on, until all the visits have been accounted for. Thus, at each stage, registrations are linked with others from earlier visits if they are close enough, and in this way separate clusters of points are built up. Such clusters containing at least three registrations from different visits (where nine or more visits were made) are taken to correspond to territories. The underlying ornithological justification for the method is that if a bird is repeatedly recorded in the same part of a plot, then this suggests that a territory may exist there. For the detailed mathematical properties of the method and associated statistical analyses, see North (1977) (reference 2).

5. General properties of the new method

The defining rules for this new method determine a unique set of clusters, for a particular defining distance, from any set of registrations. The method is therefore completely objective. It is ideally suited to use on a computer and requires as data input only a list of position and time (i.e. visit number) co-ordinates of the registrations. These can be simply entered on to tape by positioning a digitizer (an electronic pen) over each registration in turn on a covering grid. A FORTRAN programme has been written to carry out the procedure, and this produces a listing of the registrations and their grouping into clusters, both of which are also plotted in graphical output.

6. Results

Applications of the method have been successful for a number of species, notably Blackbird and Chaffinch, in the sense that good agreement has been obtained with the territory number estimates from the original analysis. This has been achieved by choosing from the sequence of defining distances d_i ($i = 1, 2, 3, \ldots$) that one which produces a result similar to that of the original analysis. Estimates agreed well despite the fact that the descriptive information had been ignored by the new method. It should be noted that the actual grouping of registrations within clusters may be somewhat different from that within B.T.O. territories, even when the numerical agreement is good, though certainly the correspondence between clusters and territories is strong (see Figures 1 and 2). We also note in Figures 1 and 2 that the large territory in the bottom left-hand corner would be located by the cluster method at a larger defining distance when the H, which is already linked with the S's, also links with A. An analysis of the differences, where they occur, between the results of the two methods has enabled an investigation to be made into the possible effects of some practical problems; one of these is the missing of records by observers because of, for example, depressed song in poor weather conditions, while another is the partial coverage of plots on some visits. For the first problem, one examines the effect on the number of territory clusters of permuting the visit letters. It is found that the missing of registrations, particularly on early visits, can lead to underestimates of the territory numbers. The second problem is well illustrated by a simple analysis, which involves carrying out χ^2 tests for contingency tables (for details, see reference 2).

7. Plans for future development

A project is planned in which the new method will be expanded further. It is likely that some of the descriptive information will have to be re-introduced if good results are to be obtained for all species, and the clustering procedure will have to be appropriately adapted. For example, the knowledge that two registrations on the same visit refer to the same male singing from two different positions, or to two different males singing simultaneously, can be very valuable when deciding how to cluster registrations in complicated cases. Which information to include, and how much, may well vary with species. Also, a method is needed for deciding how close registrations should be in order to be linked; or some alternative procedure may need to be devised which does not depend on the choice of a single clustering defining distance (which will be a function of species, habitat and population level). It is envisaged that eventually the method will be much more widely applicable than to just a few species and one broad habitat type, while retaining the desirable properties of objectivity and ease of use on a computer, which have already been achieved.

References

1. B. Everitt, *Cluster Analysis*, (Heinemann, London, 1974).
2. P. M. North, A novel clustering method for estimating numbers of bird territories, *Applied Statistics* **26** (1977), 149–155.

A Calculus Paradox

P. SHIU
University of Technology, Loughborough

1. Improper integrals

Let the continuous function f be such that $f(x) > 0$ for all $x \geqslant 1$. If $X > 1$, then the integral $\int_1^X f(x)dx$ exists and represents the area between the curve $y = f(x)$, the ordinates $x = 1$ and $x = X$, and the x-axis. Since $f(x) > 0$, it is clear that this area (or the integral) increases with X; it will therefore tend to a definite number, or to ∞, as $X \to \infty$. By the *improper integral* $\int_1^\infty f(x)dx$ we mean

$$\lim_{X\to\infty} \int_1^X f(x)dx$$

when this limit exists, and ∞ when the limit does not exist. The improper integral therefore represents the area of the region between the curve $y = f(x)$ $(x \geqslant 1)$, the ordinate $x = 1$, and the x-axis. At first sight it might appear that the condition $f(x) \to 0$ as $x \to \infty$ is sufficient to make the improper integral finite. However, this is not so. For consider the region R to the right of the ordinate $x = 1$ and lying between the x-axis and the rectangular hyperbola $y = 1/x$ $(x \geqslant 1)$. The area of R is infinite even though $1/x \to 0$ as $x \to \infty$. This is because

$$\int_1^X \frac{1}{x}\,dx = \log X \to \infty \quad \text{as } X \to \infty.$$

On the other hand, if the curve $y = 1/x$ is replaced by the curve $y = 1/x^2$, the area of the corresponding region is finite, since

$$\int_1^\infty \frac{1}{x^2}\,dx = \lim_{X\to\infty} \int_1^X \frac{1}{x^2}\,dx = \lim_{X\to\infty}\left(1 - \frac{1}{X}\right) = 1. \tag{1}$$

The essential difference between these two examples is that, even though both integrands tend to 0 as $x \to \infty$, the first becomes small rather slowly, while the second becomes small fast enough to give a region with a finite area.

Suppose next that $f(x)$ is defined for $0 < x \leqslant 1$, that f is continuous and positive, but that $f(x) \to \infty$ as $x \to 0$. We can then define $\int_0^1 f(x)dx$ as another kind of improper integral as follows. If $0 < \varepsilon < 1$, $\int_\varepsilon^1 f(x)dx$ exists as an ordinary integral and increases as ε decreases. We therefore define the improper integral $\int_0^1 f(x)dx$ as

$$\lim_{\varepsilon\to 0} \int_\varepsilon^1 f(x)dx$$

when this limit exists, and as ∞ when the limit does not exist. This improper integral represents the area of the region between the two axes, the ordinate $x = 1$, and the curve $y = f(x)$ $(0 < x \leqslant 1)$. Now consider the region between the axes, the ordinate $x = 1$, and the rectangular hyperbola $y = 1/x$ $(0 < x \leqslant 1)$. This region has infinite area since

$$\int_\varepsilon^1 \frac{1}{x}\,dx = -\log \varepsilon = \log\frac{1}{\varepsilon} \to \infty \quad \text{as } \varepsilon \to 0.$$

Note that this agrees with the result in our first example since, as we can see in Figure 1, when the present region is reflected in the line $y = x$, it coincides with our first region R except that it has an extra unit square.

2. The paradox

It is well known that if the region between the curve $y = f(x)$ and the x-axis is revolved about the x-axis, then the volume of the solid of revolution is given by $\pi \int y^2 dx$, and it does not matter whether this is an ordinary or an improper integral.

Consider now the region R in our first example. We proved that the area of R is infinite and this would indicate that we cannot paint R using only a finite amount

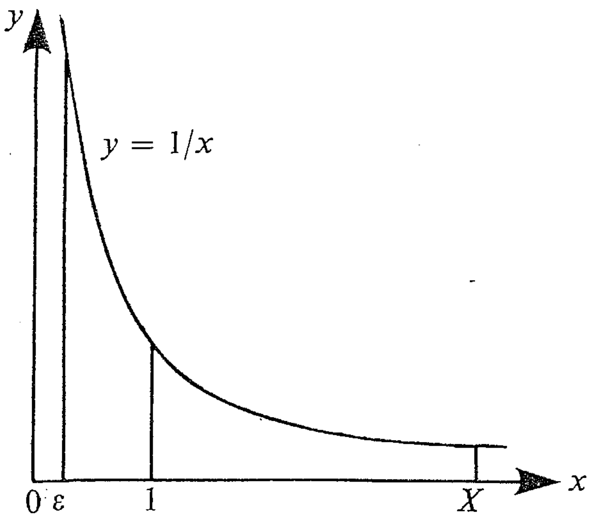


Figure 1

of paint. However, if we consider its solid of revolution S we see that the volume of S is given by

$$\pi \int_1^\infty \frac{dx}{x^2} = \pi,$$

according to (1). Since the volume is finite, we can fill the space that S occupies with paint. We note, however, that, since the region R is actually part of a cross-section of the solid S, it follows that the region R must be covered with paint after all. The paradox is: can we or can we not paint the region R?

3. Relation between area and volume of revolution

Many students, when first confronted with this paradox, suspect the mathematics involved here. It is difficult to accept that a region with infinite area can have a solid of revolution which is finite in volume. It should be pointed out here that the mathematics we used is quite correct. The main reason why the region R can give a finite volume of revolution is that most of the region R clings to the x-axis of revolution, so that it generates very little volume. On the other hand, we can show that a region with a finite area can have an infinite volume of revolution. Consider, for example, the region between the axes, the ordinate $x = 1$, and the curve $y = 1/\sqrt{x}$ $(0 < x \leqslant 1)$. The area of this region is

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim_{\varepsilon \to 0} \int_\varepsilon^1 \frac{dx}{\sqrt{x}} = \lim_{\varepsilon \to 0} 2(1 - \sqrt{\varepsilon}) = 2.$$

However, the volume of the solid of revolution corresponding to this region is infinite, since

$$\pi \int_\varepsilon^1 \left(\frac{1}{\sqrt{x}}\right)^2 dx = \pi \int_\varepsilon^1 \frac{dx}{x} = \pi \log \frac{1}{\varepsilon} \to \infty \quad \text{as } \varepsilon \to 0.$$

The gist of this example is that, in contrast to the previous case, a significant proportion of the region is very far away from the x-axis of revolution. In particular, even though the region has finite area, it is an unbounded region. It is, of course, trivial that a bounded region must have a finite volume for the solid of revolution.

4. The explanation

If it is not sloppy mathematics, how then do we explain the paradox in Section 2? Can we or can we not paint a region having an infinite area with a finite amount of paint? The explanation is that, when we see the word 'painting', which means 'covering with paint', we instinctively think of a particular way of performing this task. In fact, there are two methods of painting.

Method (*a*). We must paint with a fixed uniform thickness $\delta > 0$ throughout the region.

Method (*b*). There is no requirement on uniform thickness in the paint, and we can thin out the paint as we go along.

It is clear that if we paint a region with area A using method (a), then the volume of paint required is $A\delta$, and therefore no finite amount of paint can be used to paint by this method a region with an infinite area. On the other hand, it can be shown that, using method (b), we can paint even the whole xy-plane with only a finite volume, say V, of paint. We first arrange the xy-plane as a sequence of expanding frames with unit width. That is, we put, for $k = 0, 1, 2, \ldots,$

$$F_k = \{(x, y)\colon k \leqslant |x| < k + 1, k \leqslant |y| < k + 1\}.$$

The xy-plane is now the union of these frames F_k ($k = 0, 1, 2, \ldots$); see Figure 2. Since each frame F_k is made up of $8k + 4$ unit squares, we see that the whole plane is the union of a sequence of unit squares ($S_1, S_2, \ldots$), say. We now paint the square S_n with $1/2^n$ of the given volume V of paint; this being possible with thickness

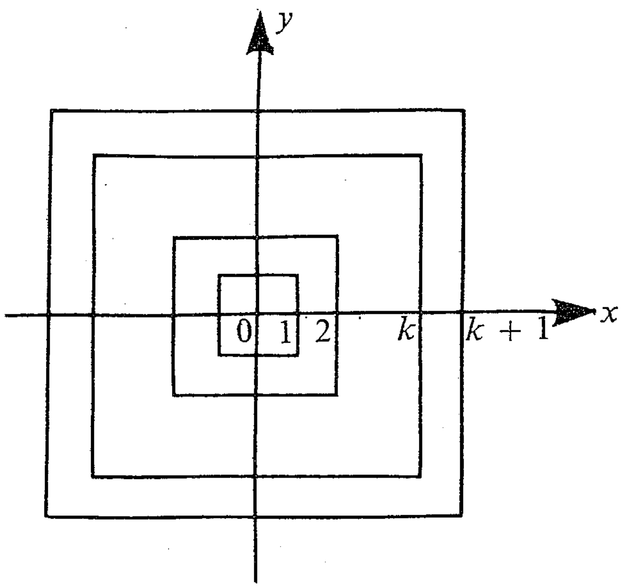


Figure 2

$V/2^n$ throughout S_n. The whole plane can be painted by this method, since the total volume of paint used in painting all the squares S_n is precisely

$$V(\tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \cdots) = V.$$

The paradox can now be explained completely. When we said in Section 2 that the region R, being part of a cross-section of its solid of revolution, is covered with paint, this type of painting is with method (b). Indeed, it is easily seen that, when the solid of revolution is filled with paint, the thickness at a point (x, y) of the region R (so that $0 \leqslant y \leqslant 1/x \leqslant 1$) is at most $2/x$. This shows that the paint is being thinned out to an arbitrarily small thickness as $x \to \infty$.

Animal Mathematics

W. D. COLLINS
University of Sheffield

Why can a sparrow fly but not an ostrich? Could some prehistoric dinosaur or brontosaurus have been very much bigger than an elephant? Why does a running mouse take more steps per minute than a horse? These questions don't seem to have much to do with mathematics, but in fact we can use some simple mathematical ideas of proportion and similarity to give plausible answers to these and other questions concerned with the effects of size on the shape of animals and how they perform certain actions such as running, jumping and flying.

A sphere of radius R has volume $\frac{4}{3}\pi R^3$ and surface area $4\pi R^2$, so that the ratio of its volume to its surface area is proportional to R. To show that this simple result can have consequences for animals and insects we'll suppose there is a group of animals, all of which are spherical in shape and made of exactly the same materials, the animals however having different radii. We can call them spheruscs. An animal loses moisture by perspiration through the skin, so the loss of moisture can be taken as proportional to its skin area. The moisture content of an animal is, however, proportional to its volume. For a spherusc of radius R this means its loss of moisture is proportional to R^2 and its moisture content to R^3. The ratio of the loss of moisture to the moisture content is then proportional to $1/R$, a ratio which increases as the size of spheruscs decreases. This larger ratio for a small spherusc means it could perspire excessively, possibly dehydrating itself. A very small spherusc is thus an impossibility unless it has a less porous skin than a large one, a skin which is hardened or thickened to cut down the loss of moisture. Whilst real animals are not spherical in shape, a version of our argument does apply to them and could explain the nature of the skins of many small land animals and insects.

Besides losing moisture animals also lose heat through their skins. A warm-blooded animal keeps its body temperature constant, so that the heat it loses through its surface must equal the heat it can produce in its body. For a warm-blooded spherusc the heat loss is proportional to R^2 and the heat produced to its volume and so to R^3. The ratio of the heat lost to the heat produced increases as the size of the spherusc decreases, so that a small spherusc must produce more heat per unit volume than a large one in order to keep up with surface loss. Since the heat an animal produces is related to the food it consumes, this extra heat production means that a small spherusc must consume more food in proportion to its weight than a large one. A man eats a fiftieth part of his own weight of food daily but a mouse half its own weight. A warm-blooded animal much smaller than a mouse becomes an impossibility, since it could neither find nor digest sufficient food to keep its temperature constant. No animals or birds are as small as the smallest frogs or fishes which are cold-blooded.

It was argued at one time that the disadvantage of small size is all the greater when loss of heat is accelerated by conduction as in the Arctic or by convection as in the seas. This was thought to explain why the polar regions are the home of large birds like penguins but not small ones, why bears but not mice live through Arctic winters and why there are no small animals in the seas. Bergmann's law, as this argument came to be known, was even used to explain why Eskimos are short and squat and desert people such as the Australian Aborigine and the Tuareg of the Sahara are tall and thin. Most biologists today however do not accept these applications of the argument. Small animals do survive Arctic winters by insulating themselves, and in some species size increases as we go from the Arctic to the Tropics.

In *Gulliver's Travels* Jonathan Swift has the Emperor of Lilliput allow Gulliver meat and drink sufficient to support 1724 Lilliputians. When Gulliver enquired how this figure was arrived at, he was told that since he was twelve times as tall as the Lilliputians, it was concluded from the similarity of their bodies that his must be equivalent to at least 1724 of theirs and so he would require as much food as was necessary to support that number of Lilliputians. Gulliver thought this was very clever of the Lilliputians, though we aren't told why the figure was not 1728. Swift, who wouldn't have known of the heat argument, which was only put forward in the nineteenth century, doesn't tell us if Gulliver got through all this food and drink. You might like to work out the number of Lilliputian portions the heat argument suggests Gulliver needed and see it the Lilliputians were as clever as Gulliver thought.

Changes in size can also make necessary changes of shape and proportion. We need only think of ourselves as we've grown from childhood to see that this is so. The ratio of the height of an adult man to the height of his head is about 7·5 but for a baby only 4. This hasn't always been appreciated by some well-known painters, since many art galleries display on their walls babies painted by old masters who have given them the adult proportions of body and head. We'll find it useful to imagine a group of animals, all of which are geometrically similar, differing from each other only in size so that each animal is an exact scale replica of any other. We'll suppose also that corresponding parts of the animals are made of the same materials. Such animals we call isometric. The spheruscs we invented are imaginary examples of such animals, but whole ranges of birds, sea animals and four-legged animals are roughly isometric. In fact, it is a group of sea animals, the porpoises, dolphins and whales, which are the most nearly isometric of all groups of animals even though there is a 5000-fold variation in weight over the group. Isometric animals can be described by a single representative length, for example, the total height or the length of a particular bone. If one of two isometric animals is k times as high as the other, then it is k times as long and k times as broad and any bone is k times as long and k times as thick as the corresponding bone in the other. The surface area of the first is k^2 times the surface area of the second and, since the animals are constructed of the same materials, the weight of the first is k^3 times the weight of the second. (See Figure 1.)

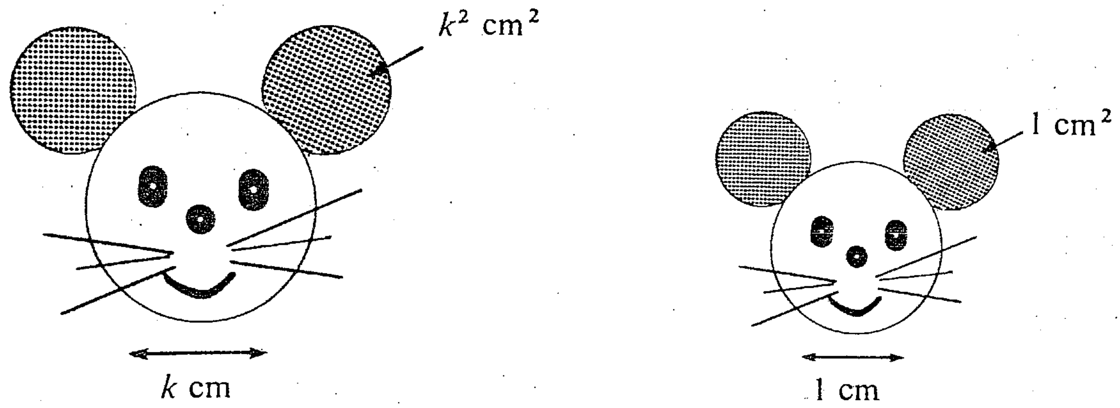


Figure 1

Changes in size mean that for four-legged animals there must be changes in the proportions of their legs. An elephant does not have legs as thin as those of a deer. The legs of an animal must be sufficiently strong to support its weight, otherwise they would bend or snap. The ability of the legs to support the animal's weight depends on how large the stresses set up in them become. Stress is force per unit area, and a bone under compression or bending breaks when the stresses in it reach critical values which are independent of the size of the bone. In the same way we can tear a piece of paper because the paper is unable to withstand the stresses applied to it where it tears, and this doesn't depend on the size of the piece of paper. For an isometric animal the area of cross-section of a bone is proportional to L^2, where L is some typical length of the animal. The forces in the legs are proportional to the animal's weight and hence to L^3, so that the stresses in the bones of the legs are proportional to L. This means that there must be an upper limit to the size of animals in a group of isometric animals, otherwise, as L increases, the critical values of the stresses are exceeded. Very large land animals such as elephants are unlikely to be isometric to smaller ones. The legs of large animals must be thickened or strengthened in some way, and the animals must stand with their legs straight rather than flexed so as not to set up too high bending stresses. Even so, there must come an upper limit to the size of animals which can stand on dry land. Sea animals do not have this problem, since their weight is supported by buoyancy. Large whales weigh over 130 tons and are very much larger than elephants, which weigh about 7 tons.

Whales can't walk but elephants can, so the upper limit to the size of land animals presumably comes somewhere in between. Estimates from skeletons suggest that some prehistoric dinosaurs were very much larger than elephants, possibly weighing as much as 80 tons. Dinosaurs as large as this may however have spent their life wading with much of their weight supported by water. There is evidence from fossil footprints however that dinosaurs weighing about 30 tons did walk on land.

When animals move, we can observe that large animals need more time for a given movement than smaller ones. For example, large birds make fewer beats of their wings per second than smaller ones, whilst a running mouse takes more steps per minute than an elephant or a horse. Ideas of proportion and similarity indicate why this should be so.

When an animal runs, those of its muscles which come into play do work which is used mainly to overcome the inertia of the moving limbs. The work done in a single contraction of a muscle is the force exerted by the muscle times the distance it contracts. For an isometric animal with typical length L the force exerted by the muscle can be taken to be proportional to its cross-sectional area and so to L^2, whilst the distance the muscle contracts is proportional to L. The work done by the muscle in a single contraction is then proportional to L^3. If we suppose this work contributes to giving a velocity V to a part of the animal's body in motion, say, a foot, then the gain in kinetic energy of this part is $\frac{1}{2}mV^2$, where m is the mass of the part, and is proportional to L^3V^2. Since the work done equals the gain in kinetic energy, this gives $L^3 \propto L^3V^2$, so that V does not vary with the size of the animal. Further, the time taken for a muscle to contract is proportional to L/V and hence to L, so that the frequency of contraction of the muscles is proportional to L^{-1}. The power available from the muscles is the rate at which they do work and equals the work done in each contraction of the muscles times the frequency of contraction. It is thus proportional to L^2.

Since isometric animals can accelerate corresponding parts of their bodies to the same velocity, they can run on the flat at the same speed irrespective of size. Greyhounds, whippets and racehorses, for example, have roughly the same top speeds even though greyhounds have legs half as long again as those of whippets and horses have legs longer than either.

That the velocity V is independent of size can also account for the differences in the numbers of steps taken by running animals. The distance a foot of an isometric animal can move relative to its body is proportional to the length of a leg and so to L for an animal of typical length L. Since the foot moves this distance at the same velocity for animals of different sizes, the time required to take a step is proportional to L, so the number of steps per minute is proportional to L^{-1}. Thus small animals take more steps per minute than large ones.

Running uphill, an animal must do work in raising its centre of gravity as well as in overcoming the inertia of its moving limbs. If an isometric animal of length L runs uphill with a vertical component U of velocity, then work must be done against gravity at a rate proportional to L^3U. If the power available from the muscles is proportional to L^2, then the vertical velocity U is proportional to L^{-1}. Speed up a hill should be inversely proportional to size. This can be observed in that, for example, a dog can race up a hill which a horse can only take at a walk.

When an animal jumps vertically, it does work against gravity. The animal first crouches in readiness to jump, stretches itself to an erect position in which its feet are just about to leave the ground, and finally takes off into free flight. (See Figure 2.) During the stretching part of the jump the animal accelerates itself from rest by

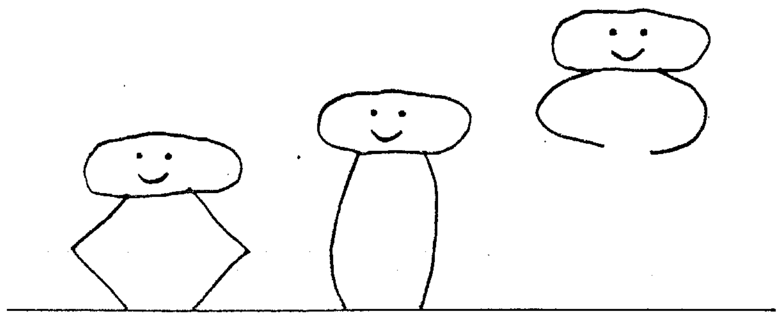


Figure 2

exerting a force F on the ground. If this force is constant, then the acceleration a of the animal is F/M, where M is its mass. If its centre of gravity moves a distance S upwards from the crouching position to the take-off position, then the velocity V with which it takes off is found from

$$V^2 = 2aS = 2FS/M.$$

When it is in free flight, the animal moves under gravity only. If its centre of gravity moves a distance H upwards during the free flight, then

$$0 = V^2 - 2gH,$$

so that

$$H = \frac{V^2}{2g} = \frac{FS}{Mg}.$$

We can assume that the force F is proportional to the cross-sectional area of the muscles producing it, so that for an isometric animal $F \propto L^2$. Also $M \propto L^3$ and $S \propto L$, so that V and H are both independent of L. An isometric animal displaces its centre of gravity a height independent of size when in free flight in a vertical jump. Observations do in fact suggest that a small jerboa can jump as high as a large kangaroo. Applying a similar argument to long jumps from a standing position, we again conclude that the distance jumped by an isometric animal is independent of size and this is borne out by such observations as there are.

How big a bird can fly? Birds up to the size of a bustard (about 12 kg) can fly, but the largest birds including the ostrich and the emperor penguin cannot. Ideas of proportion and similarity indicate why flying becomes more difficult for large birds than for small ones. We'll estimate the power a bird requires if it is to fly and the power it can produce and show that the larger the bird the more difficulty it has in producing the power required to fly. To fly horizontally at a steady speed a bird needs to produce an upward force or lift to overcome its weight and a forward force or thrust to balance the drag on it due to air resistance. A bird produces lift through the flapping of its wings. Air is deflected downwards by the wings, the bird so creating a downward jet of air. Newton's Second Law tells us that the momentum of the jet per unit time equals the lift generated. If the density of a stream of air is ρ and its velocity U, then in unit time a mass ρU of air passes through unit area perpendicular to the direction of flow, and the momentum of this mass is ρU^2. A wing of area A moving with velocity U thus causes a change of momentum of the order of $\rho U^2 A$ in unit time, so that the lift produced is proportional to $\rho U^2 A$. Assuming that U is proportional to the forward speed V of the bird, this rather crude argument indicates that the lift on the bird is $\rho V^2 AC$, where the coefficient C is independent of the size of the bird and is the same for a group of isometric birds. Since the lift balances the weight W of the bird, $W = \rho V^2 AC$, and so $W/A = \rho CV^2$. For an isometric bird of typical length L we have $W \propto L^3$ and $A \propto L^2$, so that $V \propto (W/A)^{1/2}$ gives $V \propto L^{1/2}$. This indicates that larger birds might be expected to achieve greater speeds in forward flight than smaller ones, this being borne out by observation.

The power P_r the bird requires to fly at a speed V is TV, where T is the forward force or thrust balancing the drag. Experimental evidence suggests that, for the range of speeds at which birds fly, the drag produced by air resistance is proportional to V^2L^2, so that $T \propto L^3$ and hence the power P_r required is proportional to $L^{7/2}$.

We now need to estimate the power P_g the bird can actually produce by flapping its wings. This power is provided by muscles, and so we can say that P_g equals the mass m of these muscles times the power available from each unit mass of these muscles. This latter power equals the work Q done in one contraction by each unit mass of muscle times the flapping frequency. The work Q we assume to be the same for all isometric birds. The flapping frequency f decreases with size, so that bigger birds flap their wings more slowly than smaller ones. The maximum flapping frequency of a wing is determined by the force needed to give angular acceleration to

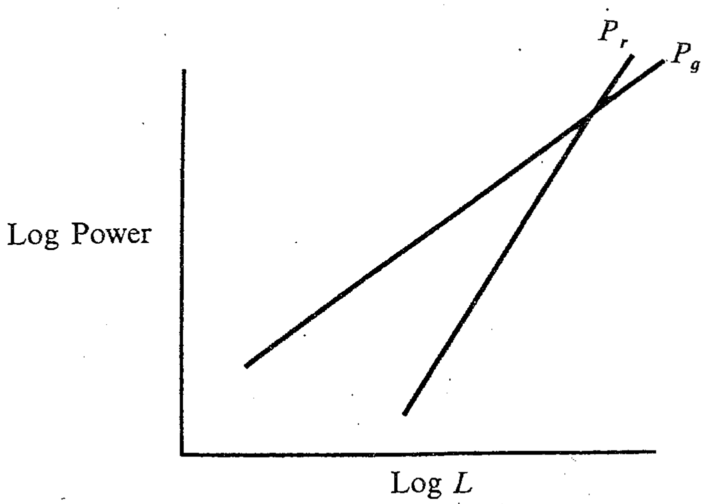


Figure 3

the wing at the top and bottom of a stroke. Scale arguments then show that $f \propto L^{-1}$. Since $m \propto L^3$, the power $P_g \propto L^2$.

If we plot the logarithms of P_r and P_g against log L (Figure 3), then we obtain straight lines of gradients 3·5 and 2 respectively. As L increases, a value is reached at which the two straight lines cross and P_r exceeds P_g for large values of L. This means that large birds cannot produce sufficient power to be able to fly, so our arguments suggest that there is a definite upper limit to the size and weight of flying birds, though they aren't able to tell us what these limits are. However, the Kori bustard, the white pelican, the mute swan and the California condor all weigh about 12 kg and are amongst the largest birds that can fly. The Kori bustard only flies for short distances and then very seldom. Some large birds such as albatrosses and vultures, which weigh about 7 to 10 kg, stay airborne for quite long periods but not in self-powered flight. Albatrosses soar in the currents of rising air produced when wind is deflected upwards by cliffs, and vultures do so in the thermals, columns of rotating air, produced on the East African plains. One consequence of our arguments is that man is too heavy to fly by self-powered wings, however large, something that wasn't realised by several of the early pioneers of flight.

There are several more examples of proportion and similarity giving us insight into the shape of animals and the performance of movements. If you would like to explore further, there are some very readable books and papers which are listed below. Except for the one by D'Arcy Thompson the books are all quite short.

References

R. M. Alexander, *Size and Shape*. Institute of Biology Studies in Biology No. 29 (Edward Arnold, London, 1971).

A. V. Hill, The dimensions of animals and their muscular dynamics. *Science Progress* **38** (1950) 209–230.

E. L. Offenbacher, Physics and the vertical jump. *Amer. J. Phys.* **38** (1970), 829–836.

C. Pennycuick, *Animal Flight*. Institute of Biology Studies in Biology No. 33 (Edward Arnold, London, 1972).

J. M. Smith, *Mathematical Ideas in Biology* (Cambridge University Press, 1968).

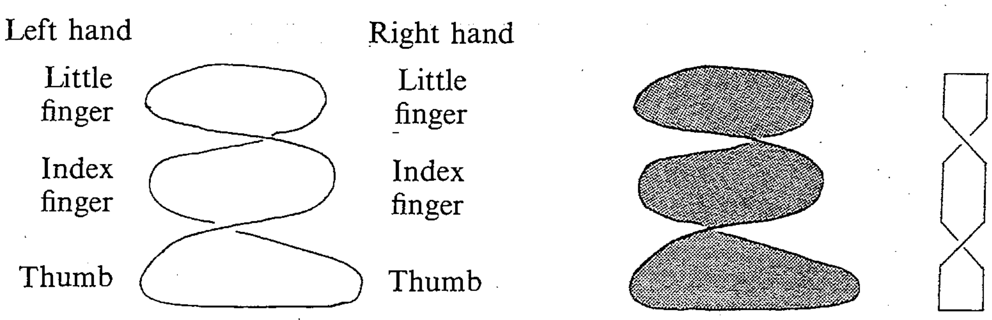
D'Arcy W. Thompson, *Growth and Form*, Abridged Edition (Cambridge University Press, 1966).

Dear Editor,

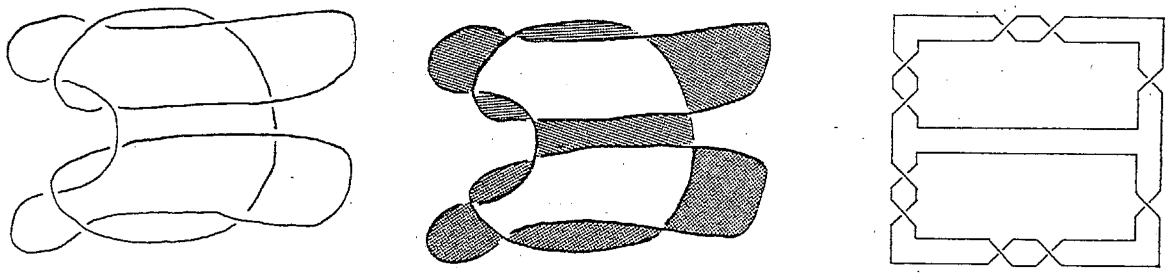
String figures

I was very interested in Dr Maddison's approach to string figures (Volume 9, Number 1, pp. 20–25) and I thought that your readers might wish to know of another possible way in which this subject might be treated. My suggestion is that a paper model can also be used to obtain a sequence of operations which leads to a given string figure.

Given a particular string figure, here is the method for constructing the paper model. The steps are illustrated in Figure 1.



(a) Navajo Opening



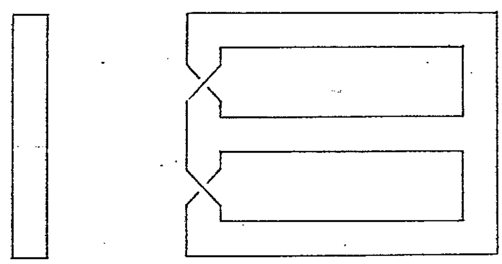
(b) The Coconut

Figure 1

1. Arrange the string figure so that all crossings are clearly separated, and make a drawing of this figure as it appears viewed from above.
2. The lines of the drawing divide the plane into a finite number of regions. There is an (infinite) outside region, which is left blank, and the other regions are shaded or left blank so as to give a chess board effect; thus, when two regions are adjacent, one is left blank and the other is shaded. It may be shown that this operation can always be carried out in a unique manner.
3. The shaded regions now represent the sides of paper (or rubber) strips with each crossing in the drawing corresponding to a half twist in a strip. Shaded regions bounded by n lines, where $n > 2$, become junctions of n strips. Intuition suggests that the construction of this paper model is always possible, and there is, in fact, a proof of this assertion.

The *edge* of the paper model, which may clearly be thought of as a string figure, can be shown to be equivalent to the original string figure. To define the notion of equivalence we consider two string figures, A and B, say. We take an arbitrary point on A and proceed along the string in an arbitrary direction, noting the sequence of what to an observer from above would appear to be the over- and under-crossings. If we can find a point on B so that, proceeding along a suitable direction, we obtain the same sequence of over- and under-crossings, then we say that A and B are equivalent. (This definition could, for some purposes, be regarded as too restrictive; for it distinguishes between two string figures in which over- and under-crossings are simply interchanged.)

The procedure is now to 'simplify' the model by cancelling twists in opposite directions. Figure 2 shows the simplified versions of the models in Figure 1. Each simplification corresponds to the reverse of two hand manipulations which could have been used in the construction of the string figure. While at first glance this method may appear to differ from Dr Maddison's, a closer look will show that the same simplifying techniques are used, namely the removal of twists, the uncrossing of loops that lie across each other, and the turning over of certain parts of the design.



(a) (b)

Figure 2

It is important to note that reconstructions of string figures, obtained either by Dr Maddison's method or from the paper model, are not unique. For instance The Coconut (Figure 1(b)), the string figure analysed by Dr Maddison, has at least six constructions. Thus, if the first step in breaking down the design is to release the right hand, then $FR5$ and $FR1$ drop out of Dr Maddison's formula (1) and there is a good deal of immediate cancellation. One can go on from here to obtain a good solution which is entirely different from the one described by Dr Maddison.

If there are several ways of making a string figure, it is impossible to say which of them is most likely to have been used unless one has studied the techniques and styles of the area from which the figure derives. Evidently Eskimos and Cape Bedford natives might be expected to employ different methods. However, in the case of The Coconut, Pacific Islands techniques would, I believe, still permit at least three likely solutions.

Again, the techniques used by Dr Maddison and myself do not necessarily pick out elegant constructions. The lifting of loops by one hand off another, and the mechanical twisting or untwisting of loops are all regarded as inelegant. Another difficulty that will not yield to the present methods is that many string designs work up in complexity until about three or four moves before the end and then simplify. Any analysis of the final pattern only is unlikely to discover this.

I hope that I have said sufficient to indicate the limitations of a purely mathematical approach to the problem of string figures, and to show that a great deal of work on the subject still remains to be done.

Yours sincerely,

PHILIP D. NOBLE

(St Cuthbert's House, Cambuslang, Lanarkshire)

Dear Editor,

Latin squares and magic squares

After I had read the article on magic squares by A. D. Misra (reference 1), a method of constructing orthogonal Latin squares came to my mind. This method of construction (which will be described later) is quite different from the one that appears in most current combinatorial mathematics books and which involves the Galois fields (see, for example reference 2). With the aid of this construction of orthogonal Latin squares, I managed to find a method of constructing diabolic magic squares of order $n \geqslant 5$, where n is a prime.

We notice that a Latin square of order n is an $n \times n$ array of the elements of a set $X = \{x_1, x_2, \ldots, x_n\}$ of n elements such that in any row and any column, each element of the set X occurs once and only once. So for example, let $X = \{A, B, C, D\}$; then

$$\begin{matrix} A & B & C & D \\ C & D & B & A \\ B & A & D & C \\ D & C & A & B \end{matrix}$$

is a Latin square of order 4. Thus all the matrices $A_{n,p}$, $B_{n,p}$ with $2 \leqslant p \leqslant n$ and $(n, p-1) = 1$ as constructed by A. D. Misra (reference 1 pp. 53–54) are by definition Latin squares.

A pair of Latin squares of order n is said to be *orthogonal* if, when one square is superimposed on the other, each position in the resulting $n \times n$ square contains an ordered pair, and each of the n^2 possible ordered pairs occurs exactly once in the array. So

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 2 & 3 \\ 2 & 3 & 4 & 5 & 1 \\ 5 & 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 1 & 2 \end{matrix} \quad \text{and} \quad \begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 1 & 2 \\ 5 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 & 1 \\ 4 & 5 & 1 & 2 & 3 \end{matrix}$$

are orthogonal since

$$\begin{matrix} (1, 1) & (2, 2) & (3, 3) & (4, 4) & (5, 5) \\ (4, 3) & (5, 4) & (1, 5) & (2, 1) & (3, 2) \\ (2, 5) & (3, 1) & (4, 2) & (5, 3) & (1, 4) \\ (5, 2) & (1, 3) & (2, 4) & (3, 5) & (4, 1) \\ (3, 4) & (4, 5) & (5, 1) & (1, 2) & (2, 3) \end{matrix}$$

consists of n^2 distinct ordered pairs ($n = 5$).

We say that a set of Latin squares is an orthogonal set if the squares are pairwise orthogonal. It can be proved (see reference 2) that there are at most $n - 1$ Latin squares in a set of orthogonal Latin squares of order n.

Theorem 1. Let $n \geqslant 3$ be a prime. Then the set of Latin squares, where each square is either $A_{n,p}$ or $B_{n,p}$ with p taking each of the values $2, 3, \ldots, n$ forms an orthogonal set with $n - 1$ elements.

Proof. Since when n is a prime greater than 2, we have $(n, s) = 1$ for $2 \leqslant s \leqslant n - 1$, therefore, by Lemma 1 (reference 1) and by the definition of a Latin square, all the matrices $A_{n,2}, A_{n,3}, \ldots, A_{n,n}$ are Latin squares.

Now let $2 \leqslant p < q \leqslant n$ and $(a_{n,p}(i,j), a_{n,q}(i,j))$ and $(a_{n,p}(k,l), a_{n,q}(k,l))$ be the (i,j)th and the (k,l)th elements of the resultant square obtained by superimposing $A_{n,q}$ on to $A_{n,p}$. Then

$$(a_{n,p}(i,j), a_{n,q}(i,j)) = (a_{n,p}(k,l), a_{n,q}(k,l))$$

$$\Rightarrow [(i-1)(p-1)+j-1](\operatorname{mod} n) = [(k-1)(p-1)+l-1](\operatorname{mod} n)$$

and

$$[(i-1)(q-1)+j-1](\operatorname{mod} n) = [(k-1)(q-1)+l-1](\operatorname{mod} n)$$

$$\Rightarrow [(i-k)(p-1)+(j-l)] \equiv 0 (\operatorname{mod} n)$$

and

$$[(i-k)(q-1)+(j-l)] \equiv 0 (\operatorname{mod} n)$$

$$\Rightarrow (q-p)(j-l) \equiv 0 (\operatorname{mod} n)$$

and

$$(q-p)(i-k) \equiv 0 (\operatorname{mod} n)$$

$$\Rightarrow j = l \text{ and } i = k \qquad (\text{since } 1 \leqslant q-p < n-1 \text{ and so } ((q-p),n) = 1).$$

Since this is true for $1 \leqslant i,j \leqslant n$, the result follows. In just the same way we see that $A_{n,p}$, $B_{n,q}$ are orthogonal and $B_{n,p}$, $B_{n,q}$ are orthogonal $(p \neq q)$.

With the aid of the above theorem and Lemmas 1, 2, 3 established by A. D. Misra, one can easily construct diabolic magic squares of order $n \geqslant 5$, where n is a prime.

Theorem 2. Let $n \geqslant 5$ be a prime, then $D_{n,p,q}$ given by

$$D_{n,p,q} = B_{n,p} + nA_{n,q}$$

is a diabolical magic square of order n, where p, q are integers such that $3 \leqslant p, q \leqslant n - 1$ and $p \neq q$.

(N.B. A diabolical magic square, also known as the nasik or pandiagonal square, of order n is a magic square not only possessing the usual properties of an nth-order (traditional) magic square, but in addition, each left-to-right and each right-to-left diagonal is also 'magic', i.e. each of the sums is also equal to the magic constant, namely $\frac{1}{2}n(n^2+1)$.)

Proof. Since $(p-1), (q-1) < n$, so $(n, p-1) = 1$ and $(n, q-1) = 1$. Thus by Lemma 1 of reference 1, each column of $A_{n,q}$ contains all the integers $0, 1, 2, \ldots, n-1$; and each column of $B_{n,p}$ contains all the integers $1, 2, 3, \ldots, n$. Hence each row- and column-sum of $D_{n,p,q}$ equals $\frac{1}{2}n(n^2+1)$.

Furthermore, since $3 \leqslant p, q \leqslant n-1$ and (n,p), (n,q), $(n,p-2)$, $(n,q-2)$ are all co-prime pairs, thus by Lemmas 2, 3 of reference 1, all the left-to-right and right-to-left diagonals of $A_{n,q}$, $B_{n,p}$ contain all the integers $(0, 1, 2, \ldots, n-1), (1, 2, \ldots, n)$ respectively. So every left-to-right and right-to-left diagonal sum of $D_{n,p,q}$ equals $\frac{1}{2}n(n^2+1)$.

Finally, since $p \neq q$ and $(n, p-1) = 1$, $(n, q-1) = 1$, it follows from Theorem 1 that $A_{n,q}$ and $B_{n,p}$ are orthogonal Latin squares. Hence all the entries of $D_{n,p,q}$ are distinct. The least integer which occurs in $D_{n,p,q}$ is $1 + n(0) = 1$, while the largest integer is $n + n(n-1) = n^2$. Thus $D_{n,p,q}$, as constructed above where $n \geqslant 5$ is a prime and p, q are integers such that $3 \leqslant p, q \leqslant n-1$ and $p \neq q$, is a diabolic magic square of order n. For example

$$D_{5,4,3} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 2 & 3 \\ 2 & 3 & 4 & 5 & 1 \\ 5 & 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 1 & 2 \end{bmatrix} + 5 \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 0 & 1 \\ 4 & 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 & 0 \\ 3 & 4 & 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 7 & 13 & 19 & 25 \\ 14 & 20 & 21 & 2 & 8 \\ 22 & 3 & 9 & 15 & 16 \\ 10 & 11 & 17 & 23 & 4 \\ 18 & 24 & 5 & 6 & 12 \end{bmatrix}.$$

References
1. A. D. Misra, Magic squares, *Math. Spectrum* **8** (1975/76), 53–60.
2. H. J. Ryser, *Combinatorial Mathematics* (Wiley, New York, 1963).

Yours sincerely,

K. H. YIM
(University of Manchester Institute of Science and Technology)

Problems and Solutions

Sixth formers and students are invited to submit solutions to some or all of the problems below: the most attractive solutions will be published in subsequent issues. When writing to the Editorial Office, please state your full name and the postal address of your school, college or university.

Problems

10.4. (Submitted by B. G. Eke, University of Sheffield.) The integers $a_1, a_2, \ldots, a_7$ are rearranged to give $b_1, b_2, \ldots, b_7$. Show that

$$(a_1 - b_1)(a_2 - b_2)\ldots(a_7 - b_7)$$

is even.

10.5. The real polynomials $f_1(x), \ldots, f_{n-1}(x)$, $g(x)(n > 1)$ are such that

$$f_1(x^n) + xf_2(x^n) + \cdots + x^{n-2}f_{n-1}(x^n) = (1 + x + x^2 + \cdots + x^{n-1})g(x).$$

Show that $f_1(x), \ldots, f_{n-1}(x)$ all have $x - 1$ as a factor.

10.6. (Submitted by L. Mirsky, University of Sheffield.) Let (a, b) denote the highest common factor of a and b. For any positive integers a, b, m, n with $(a, b) = 1$, show that

$$(a^m - b^m, a^n - b^n) = a^{(m,n)} - b^{(m,n)}.$$

Solutions to Problems in Volume 9, Number 3

9.7. The seven small circles in the figure all have unit radius. Find a quadratic equation, one of whose roots is the radius of the large circles. What is the geometrical significance of the other root of the quadratic equation?

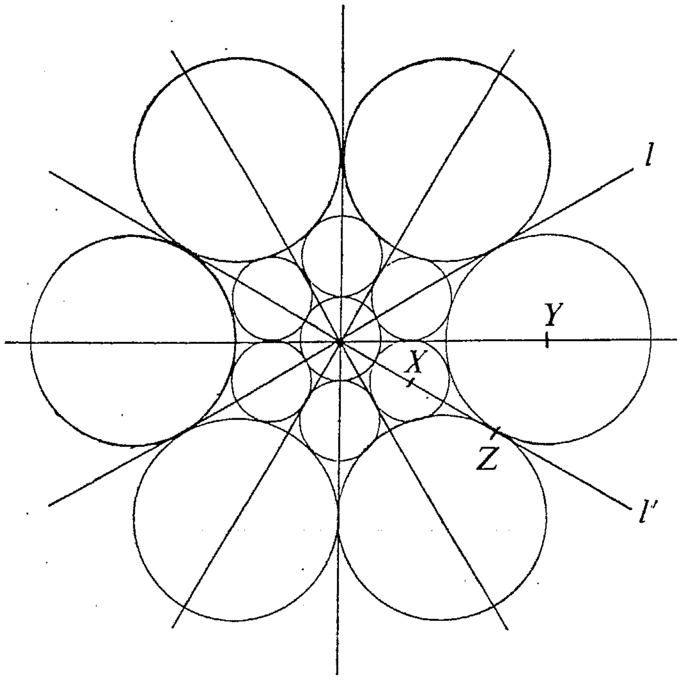


Figure 1

Solution by Thomas Lawton (Charterhouse)

Let r denote the radius of the large circles. If we consider the right-angled triangle XYZ (see the figure), we have

$$(r+1)^2 = (r\sqrt{3}-2)^2 + r^2,$$

which is a quadratic equation one of whose roots is the radius of the large circles. The other root is the radius of the circle which touches l, l' and the central circle of the diagram, and which lies inside this central circle.

Also solved by John Ramsden (University of Aston).

9.8. Let S_1 denote the sequence of positive integers $1, 2, 3, 4, 5, 6, \ldots$, and define the sequence S_{n+1} in terms of S_n by adding 1 to those integers in S_n which are divisible by n. (Thus, for example, S_2 is $2, 3, 4, 5, 6, 7, \ldots$, S_3 is $3, 3, 5, 5, 7, 7, \ldots$). Determine those integers n with the property that the first $n-1$ integers in S_n are n.

Solution

We note first that, for each r, S_r is a non-decreasing sequence. *Suppose first that n is prime.* The first term of S_n is n. The $(n-1)$th term of S_2 is n. Because n has no proper factors, this means that the $(n-1)$th term of S_n is also n. Thus the first $n-1$ terms in S_n are n. Also, the nth term of S_2 is $n+1$, so the nth term of S_n is greater than n. Thus prime numbers have the given property. *Now suppose that n is not prime*, and let p be its smallest prime factor. Then the $(n-1)$th term of S_2 is n and the $(n-1)$th term of S_{p+1} is $n+1$, so the $(n-1)$th term of S_n is at least $n+1$. Thus there are fewer than $n-1$ integers in S_n which are equal to n. This shows that the integers with the given property are precisely the prime numbers.

Also solved by J. Pemberton (U.M.I.S.T.)

9.9. Let

$$u_n = \left(1+\frac{1}{n}\right)^n, \qquad v_n = \left(1+\frac{1}{n}\right)^{n+1}.$$

Show that the sequence (u_n) is strictly increasing, whereas (v_n) is strictly decreasing.

Solution

We use the inequality

$$\text{geometric mean} \leqslant \text{arithmetic mean},$$

with a strict inequality when not all the numbers are equal. If we apply this inequality to a and n copies of b with $a \neq b$, we obtain

$$(ab^n)^{1/(n+1)} < \frac{a+nb}{n+1}.$$

First take $a = 1$, $b = 1 + (1/n)$. This gives

$$\left(1 + \frac{1}{n}\right)^n < \left(1 + \frac{1}{n+1}\right)^{n+1},$$

which shows that (u_n) is a strictly increasing sequence. Now take $a = 1$, $b = 1 - (1/n)$. Then

$$\left(1 - \frac{1}{n}\right)^{n/(n+1)} < \frac{n}{n+1},$$

$$\left(\frac{n}{n-1}\right)^{n/(n+1)} > \frac{n+1}{n},$$

$$\left(1 + \frac{1}{n-1}\right)^n > \left(1 + \frac{1}{n}\right)^{n+1},$$

which shows that (v_n) is a strictly decreasing sequence.

Book Reviews

The VNR Concise Encyclopedia of Mathematics. Edited by W. GELLERT, H. KÜSTNER, M. HELLWICH and H. KÄSTNER. Van Nostrand Reinhold Company Ltd, Wokingham, 1977. Pp. 760 + 56 plates. £12·10.

As its title suggests, this is a book for dipping into, though it is much more than just a reference book of formulae and theorems. The aim is to give concise descriptions of the various branches of mathematics which have emerged from antiquity to the present day, emphasizing the key ideas and their interrelationships. Quite a task!

After an introductory historical survey the text is in three main parts: elementary mathematics, steps towards higher mathematics, and brief reports on selected topics. 'Elementary' is here a technical term, referring to those branches of mathematics which existed before the main development of calculus, though these are treated in a modern way. The pace varies widely: for example the description of the use of logarithms and slide rule for basic calculations (including the mysterious 2·1.5) takes about ten pages, and is immediately followed by the development of the number system from Peano's axioms via integers, rationals, reals (defined by nested rational intervals), continued fractions, and complex numbers to de Moivre's theorem, also in ten pages. So 'elementary' is not equivalent to 'easy'.

Of the steps towards higher mathematics about half are concerned with limiting processes: series, calculus, differential equations. The remainder include logic, abstract and linear algebras, projective geometry, probability, statistics and numerical methods. There is a great deal of information here, although two fairly random items (Chebyshev polynomials and Laplace transforms) which one might expect to find, prove to be missing.

The twelve brief reports are intended to give insight into some research fields of contemporary mathematics. These are good introductory essays, more straightforward than some of the preceding chapters; a keen sixth former should be able to tackle them without being overwhelmed.

A helpful colour code (yellow for definitions and formulae, blue for examples, and pink for theorems) makes the book easy to use, despite the small type. Colour also emphasizes the main steps in complicated arguments, and makes the many fine diagrams clear and attractive. There are some minor misprints and also some oddities (for example, an extract from a Post Office Savings book illustrating simple interest shows a period which includes the transfer to decimal currency, so that entries change from £sd to £p); and there is a rather strange collection of photographs at the end which does not seem to add much. But these flaws do not significantly spoil an interesting compendium which is well worth considering for the library.

City of London School T. J. HEARD

Introductory Statistics: a Decision Map. By THAD R. HARSHBARGER. Collier MacMillan, London, 1977 (2nd edition). Pp. 596. £9·75.

'Yet another introductory statistics textbook' was my first reaction, but, having read this one, I feel that the author has succeeded in writing a book which fulfils his intention, since it should prove to be a suitable introductory text for students in the social sciences and education. This is not to say, however, that it is unsuitable for other students. Indeed, it is a book which could benefit many mathematical pupils and their teachers.

The approach used by the author is to present the reader with a scheme for classifying and solving problems. This is done for almost all the chapters and the scheme is called a decision map. The user is intended to approach the map for a given chapter with a specific problem in mind and then, by following the decision map (flowchart), a solution to the problem should be obtained. At first sight it gives the impression of being a recipe book, but closer consideration shows that it is much more than this. In fact, each time a solution is found, the author points out which relevant earlier sections should have been read so that, by following this earlier material, a much more comprehensive explanation is obtained.

The text contains very little in the way of mathematical proofs, but at the same time it does not hesitate to use mathematical symbols and, where explanations of their use are needed, these are included. There is strong emphasis on the assumptions to be satisfied before using the statistical techniques, and consideration is also given to the writing of conclusions once the calculations have been done; both of these are important to the work of the statistician.

There are, however, one or two places where improvements might be made. Histograms are rather more than just bar charts without spaces between the bars and I think students would, perhaps, find the concept of probability distributions easier after the chapter on probability rather than before it. The t, χ^2 and F distributions together with the concept of degrees of freedom appear earlier than is often the case in other books, although this early appearance produces no real advantage. The final chapters on correlation are very full, perhaps rather too full for an introductory text of this nature.

Overall, in spite of some notation different from that normally used, the book is well worth reading and succeeds in its aim. The preface can be used as a genuine guide when trying to decide upon the suitability of this textbook and I am sure that many mathematical and non-mathematical students could benefit from the approach.

Teesside Polytechnic G. E. SKIPWORTH

Probability and Statistical Inference. By ROBERT V. HOGG and ELLIOTT A. TANIS. Collier Macmillan, London, 1977. Pp. ix + 420. £11·25.

One of the central themes of modern statistics is statistical inference, whereby the statistician attempts to draw conclusions from the evidence provided by samples, this being done in terms of probability statements. Thus it would seem fitting that the authors have united a study of probability and a discussion of statistical inference in one well-presented book.

The style of the text is mathematical and its aim is to provide a book that emphasizes fundamental concepts and presents them in a logical order. This it certainly does and the book should be of interest to any student or teacher wanting to see the proofs of many of the statistical statements usually taken for granted. Only on rare occasions do the authors accept that proofs are beyond the scope of the text and even on these occasions they try to give some intuitive reasoning for the statements they make but do not prove. The material covered includes probability, distributions of both discrete and continuous type, together with a discussion of their parameters, estimation, tests of statistical hypotheses, chapters on multivariate distributions, chi-square tests and analysis of variance. The final chapter gives a brief theory of statistical inference and is intended to whet the appetites of its readers in the hope that they will want to learn more. Little attention is given to the problems of practical statistics and what reference is made to sample statistics is quickly absorbed into a mathematical background.

Overall it is a well-written text including a plentiful supply of worked examples and a large number of exercises for the student to work. Nevertheless it is probably beyond the scope of current school statistics students, with the exception of those who are very mathematically motivated. However, as courses develop, it could well find its place as a reference text on the shelves of many school libraries.

Teesside Polytechnic G. E. SKIPWORTH

Further Mathematical Diversions. By MARTIN GARDNER. Penguin Books Ltd, Harmondsworth, 1977. Pp. 256. £0·80.

This material, which first appeared in *Scientific American* in 1961–63, is now available as a Pelican book. What can one say about this book of Martin Gardner's except that it is the usual splendid mixture of puzzles, articles, paradoxes and jokes, and that there is something of interest for mathematicians of all ages and all abilities!

University of Durham H. NEILL

A Survey of Modern Algebra (fourth edition). By GARRETT BIRKHOFF and SAUNDERS MACLANE. Collier Macmillan, London, 1977. Pp. xii + 500. £10·15 hardback.

This well-known book appears for the fourth time with only very minor changes from the third edition.

Notes on Contributors

T. H. Heard read mathematics at Jesus College, Cambridge. He then taught for two years at a secondary school in Tanzania before taking up a post in 1967 at the City of London School, where he has been Head of Mathematics since 1971. The first volume of his sixth-form course *Extending Mathematics* appeared in 1974; the second volume (written with D. R. Martin) is being published early in 1978. He is spending the present academic year on secondment as B.P. Research Fellow in the Department of Engineering Science, Durham University, investigating the mathematical requirements and difficulties of engineering undergraduates, and the links between these and their school and university mathematics courses.

Hazel Perfect has taught in both schools and universities, and is at present a Senior Lecturer in Pure Mathematics in the University of Sheffield. Her principal interests are in matrix theory and in combinatorics. She is the author of two books, *Topics in Geometry* and *Topics in Algebra* (both published by Pergamon Press) which link school and university mathematics. She has contributed to *Mathematical Spectrum* on two previous occasions.

Philip North is a graduate of the University of Sussex and is at present a research student, supported by the Science Research Council, working in the Mathematical Institute at the University of Kent in collaboration with the British Trust for Ornithology. He has previously held teaching posts both in England and in the Netherlands.

Peter Shiu Man-Kit is a Lecturer in Mathematics at Loughborough University of Technology. As a teacher he is particularly interested in the problem of how to minimize the difficulties that students encounter in their transition from school- to university-type mathematics. His research interests are in analysis and the theory of numbers. Some years ago he contributed to *Mathematical Spectrum* an article on rational approximations to square roots.

W. D. Collins, a graduate of the University of London, is a Professor of Applied Mathematics in the University of Sheffield. After working at the Admiralty Research Laboratory he was a lecturer in the Universities of Newcastle and Manchester and then for ten years Professor of Mathematics in the University of Strathclyde. His research interests are in control theory and applied analysis.

Contents

ISSN 0025–5653

PRICES (*postage included*)

Prices for Volume 10 (*Issues Nos.* 1, 2, *and* 3):

Subscribers in Britain and Europe: £1·15
Subscribers overseas: £2·30 (US\$4.00; \$A3.70)
(These prices apply even if the order is placed by an agent in Britain.)
A discount of 10% is allowed on all orders for five or more copies.

Back issues:

Volume 1 is out of print. All other back issues are still available; information concerning prices may be obtained from the Editor.

Enquiries about rates, subscriptions and advertisements should be directed to:

Editor—*Mathematical Spectrum*,
Hicks Building,
The University,
Sheffield S3 7RH, England.

Printed in England by Galliard (Printers) Ltd, Great Yarmouth