Mathematical Spectrum

Volume 1 1968/9 *Number 1*

A Magazine of *Contemporary Mathematics*

Published by *Oxford University Press*

Mathematical Spectrum is a magazine for the instruction and entertainment of student mathematicians in schools, colleges, and universities. It is published by Oxford University Press on behalf of the Applied Probability Trust, a non-profit making organization 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 1 of *Mathematical Spectrum* (1968/9) will consist of two issues, the second of which will be published in the Spring of 1969.

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.

The Editorial Committee welcomes correspondence, queries, and solutions to problems. A selection of this material will be published in each issue. All correspondence about the contents of *Mathematical Spectrum* should be sent to:

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

Editorial

Early in 1967, a small group of mathematicians began to think of launching a mathematical magazine for readers in schools, colleges of education, and universities. In discussing the problems of this enterprise, the Organizing Committee felt that any decisions should be based primarily on ascertainable facts, while their belief in the need for a magazine should play only a subsidiary role.

It was therefore agreed to send, with the help of Oxford University Press, some 2000 copies of a preliminary announcement and mock-up of *Mathematical Spectrum* to 1100 institutions in Britain, the U.S.A. and Australia. The mock-up contained four short articles on the following topics:

The irrationality of π,

Meteoric dust and noctilucent clouds,

The spread of an epidemic,

G. H. Hardy and British mathematics.

These were similar in style and content to those projected for the new magazine; the choice of topics was intended to indicate the breadth of subjects to be covered in future issues. A number of problems were also stated, and their solutions were included. A questionnaire was enclosed, asking for information on the likely number of subscribers and their interest in various types of articles.

The response to the preliminary announcement was immediate and enthusiastic. Replies from students and teachers of mathematics confirmed the belief of the Organizing Committee that a mathematical magazine of the type planned would be welcome and useful. Many students wrote to us at length and gave detailed advice on the contents and level of articles; none of the writers left us in any doubt as to their interest or their wish to take part in the venture.

Answers to the questionnaire indicated that mathematical problems were universally popular, and that expository articles in various branches of mathematics would find appreciative readers. In planning the first issue of *Mathematical Spectrum*, the Editorial Committee have kept these preferences in mind.

The contents of the magazine, reflected in its name, will illustrate the whole range of modern mathematics: articles will appear on topics in pure mathematics, applied mathematics, statistics, operational research, computing science, numerical analysis, and biomathematics. These will be written by practising research workers and teachers in various fields at a level intelligible to students and also, we hope, to laymen interested in mathematics. Articles may be expository or historical; some elementary research will also be published. Information on careers and educational opportunities will appear from time to time. There will be a number of problems in each issue, with solutions appearing in the next.

Correspondence and answers to problems are welcome; a selection from readers' letters and the most attractive solutions to problems will be published in each issue of *Mathematical Spectrum*. We intend the magazine to develop through the active participation of readers: we invite your co-operation. Please tell us your interests and your ideas; send us your suggestions on any subject connected with the magazine, or, more generally, with mathematics and mathematical education. Your response will help us to make *Mathematical Spectrum* a periodical that is tuned in to the wavelength of its readers.

THE EDITORIAL COMMITTEE

An Extremal Problem in Elementary Geometry

R. J. WEBSTER
University of Sheffield

A farming community consists of several farms which are scattered over a wide area of flat country, and the greatest distance apart of any two farmhouses in this community is 1 mile. It has been decided that the community should be provided with a fall-out shelter, and that this shelter should be positioned in such a way that its distance from the farmhouses which are farthest away should be as small as possible. We are interested to see how effectively these requirements can be satisfied. If the shelter were to be placed at one of the farmhouses, then no farmhouse would be more than 1 mile from it. However, as we shall demonstrate below, it is possible to improve on this trivial construction.

As a simple illustration of the general problem, let us consider a three-farm community, where the greatest distance apart of any two of the farmhouses is 1 mile. If these three houses lie in a straight line or are the vertices of an obtuse-angled triangle, then the shelter is best situated midway between the two farmhouses which are 1 mile apart, so that no farmhouse is more than $\frac{1}{2}$ mile from the shelter. We consider now the remaining alternative, i.e., the case when the three farmhouses A, B, C, say, are the vertices of an acute-angled triangle. In this case, let A be the largest angle of the triangle ABC, so that $60° \leqslant A \leqslant 90°$ and $BC = 1$. Now the radius R of the circumcircle of the triangle ABC is given by the formula

$$R = 1/(2\sin A).$$

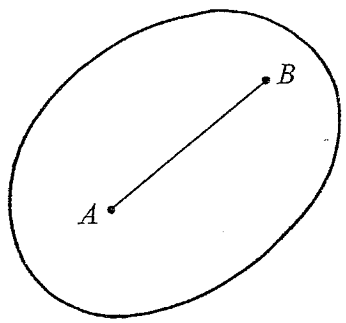
Since, moreover, $\sin 60° = \sqrt{3}/2$ and $60° \leqslant A \leqslant 90°$, we infer that

$$R = 1/(2\sin A) \leqslant 1/2(\sqrt{3}/2) = 1/\sqrt{3}.$$

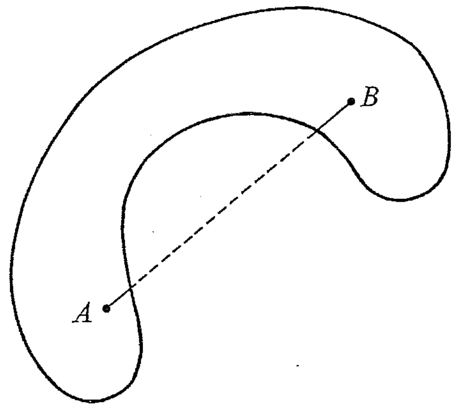
(Here we have used the fact that in the interval between 0° and 90° the sine of an angle increases with the angle.) Thus, by placing the shelter at the circumcentre of the triangle ABC, we can make sure that no farmhouse is further than $1/\sqrt{3}$ miles away from it.

Summarizing our conclusions, we see that in a three-farm community the shelter can always be placed in such a way that no farmhouse is more than $1/\sqrt{3}$ miles from it. The surprising fact is that this result is still valid for a community with *any* number of farms. This was proved by H. W. E. Jung in 1901 and (stated in abstract terms) asserts that *if a collection of n points in a plane is such that the maximum distance between any two points of the collection is* 1, *then there exists another point whose distance from none of the n given points exceeds* $1/\sqrt{3}$. The remainder of the article is devoted to a proof of this result.

The method by which we reach our conclusion depends on the idea of 'convexity'. Throughout, we assume that all figures considered here lie in some fixed plane. By a triangle, we shall understand the totality of points both on its sides and in its interior, with a similar convention applying to other figures. A *convex figure* is a special type of figure which has the property that, if two points A, B belong to it, then so does the whole segment AB. Many of the familiar figures such as the triangle, the circle, or the ellipse are instances of convex figures. The idea of convexity is illustrated in Fig. 1, (a) and (b), below. Figure (a) is



(a)



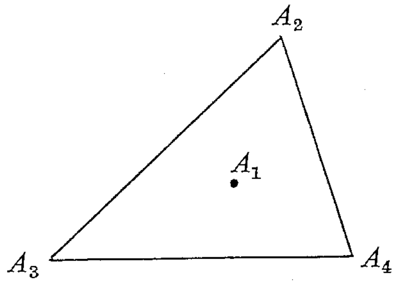
(b)

Figure 1

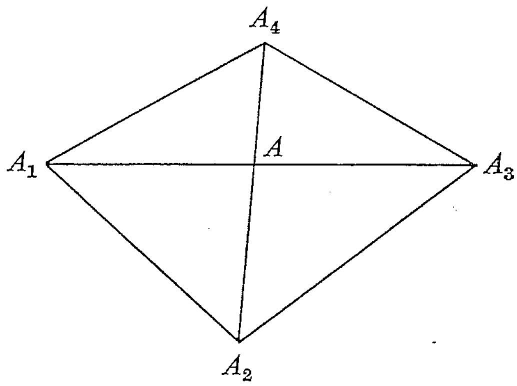
convex, whereas Figure (b) is not, since although it contains the points A and B, the segment AB does not belong to it in its entirety. It is easily verified that if A, B, C are any three points of a convex figure, then the whole triangle ABC also belongs to it. When F_1, F_2 are two figures, then we shall denote by $F_1 \cap F_2$ the figure consisting of those points which belong to both F_1 and F_2. It is almost obvious that, if F_1 and F_2 are convex figures, then so is $F_1 \cap F_2$.

The result of Jung stated a few lines earlier is equivalent to the assertion that *if a collection of n points in a plane is such that the maximum distance apart of any two points in the collection is* 1, *then all the n circles centred at the points of the collection and of radii* $1/\sqrt{3}$ *have a point in common.* We are thus led to consider the common points of circles, or more generally the common points of convex figures. The reason why we do not confine ourselves to the study of common points of circles is that the common points of a collection of circles do not form a circle, whereas the common points of a collection of convex figures do form a convex figure.

The first result we shall establish in this context is that *if four convex figures* *are such that any three of them have a common point, then all four have a common point.* Let, then, F_1, F_2, F_3, F_4 be convex figures; let A_1 be a common point of F_2, F_3, F_4; let A_2 be a common point of F_1, F_3, F_4; let A_3 be a common point of F_1, F_2, F_4; and let A_4 be a common point of F_1, F_2, F_3. Two alternatives are possible: either one of the points A_1, A_2, A_3, A_4 lies in the triangle formed by the other three or else the four points constitute a convex quadrilateral. These two alternatives are illustrated in (a) and (b) of Figure 2. Consider the first case and assume (as may be done without loss of generality) that A_1 is a point of the triangle $A_2A_3A_4$. Since F_1 is a convex figure and since A_2, A_3, A_4, are points of F_1, we infer that the whole triangle $A_2A_3A_4$ belongs to F_1. In particular, then, A_1 is a point of F_1. But we already know that A_1 is a point of each of F_2, F_3, F_4. Thus A_1 is a point common to F_1, F_2, F_3, F_4; and our result is proved in the first case.



(a)



(b)

Figure 2

Consider now the second case and assume that A_1, A_2, A_3, A_4 are the vertices of a convex quadrilateral as illustrated in (b) of Figure 2. Let the diagonals of this quadrilateral intersect in the point A as shown. Since A_1, A_3 are points of F_2 and since F_2 is convex, we infer that the whole segment A_1A_3 belongs to F_2. In particular, A belongs to F_2, and similarly A belongs to F_4. Again, using the fact that A lies on the segment A_2A_4, we infer in a similar way that A belongs to F_1 and to F_3. This shows that A is a common point of F_1, F_2, F_3, F_4, and the proof of our assertion about common points is therefore complete.

We now generalize the result just considered to an analogous result for n $(n \geqslant 3)$ convex figures. We shall, in fact, show that *if n convex figures are such that every three of them have a common point, then all n have a common point.* This result is known as Helly's theorem; it was proved by E. Helly in 1923, but earlier proofs had actually been given by J. Radon and D. König. The proof proceeds by induction with respect to n. We note that the result is obvious for $n = 3$. *Assume* that the result is valid for $n = m$ and let $F_1, F_2, \ldots, F_m, F_{m+1}$ be $m+1$ convex figures, any three of which possess a common point. Consider now the m convex figures $F_1, F_2, \ldots, F_{m-1}, F_m \cap F_{m+1}$. Any three of these figures have a common point. For example, to see that F_1, F_2, $F_m \cap F_{m+1}$ have a common point, we note that (in view of our hypothesis) any three among F_1, F_2, F_m, F_{m+1} have a common point. Hence, by the result proved a little earlier, all four figures have a common point, and this point is clearly common to F_1, F_2, $F_m \cap F_{m+1}$. Since, then, any three of the m convex figures $F_1, F_2, \ldots, F_{m-1}, F_m \cap F_{m+1}$ have a common point, it follows by the induction hypothesis that all m figures have a point in common. This point is clearly a common point of the $m+1$ figures $F_1, F_2, \ldots, F_m, F_{m+1}$. Helly's theorem is therefore proved.

We now come to Jung's result. Consider any collection of n points in a plane such that the maximum distance between any two of them is 1, and consider also the n circles which have radii $1/\sqrt{3}$ and are centred at the points of this collection. In view of what we have proved concerning the positioning of the fall-out shelter for a three-farm community, it follows that any three of these circles have a common point. Hence, by Helly's theorem, there is a point common to all n circles; and this point is at a distance not exceeding $1/\sqrt{3}$ from each of the n points of the collection. This completes the solution of our original problem, but it might also be mentioned that all the results we have described can be extended to configurations in three-dimensional space.

Where Shall We Build It?

F. BENSON

University of Southampton

An important area of Operational Research, namely the application of scientific methods to industrial and social problems, is concerned with the siting of homes, factories, or community projects in a particular geographical region. For example, a city council may be discussing a proposal to build a community centre on a new housing estate. A map of the estate is prepared showing a few alternative sites suggested by the architect. The council must now decide which of these suggestions to adopt or even, perhaps, pick a different site from any of those suggested.

The same housing estate will also be under discussion at the regional headquarters of the Electricity Board. The estate is a large one and a new substation will be needed; the board must decide where it should be built.

At the same time, a large company, with several factories in various parts of the country, has acquired a large new factory in the city. The company owns a warehouse from which goods are distributed to its customers, but because of the changed location of the new factory, the warehouse has become unsuitable. The increased trade also makes the warehouse too small. A new warehouse must be built. Where should it be?

In every case where a construction project is planned, a large part of the discussion is concerned with the design and cost of the new buildings. The building cost might depend on the choice of site, but it is unlikely to vary much in comparison with the total eventual cost of the project. What factors then decide where the building is to go? The city councillors are concerned with the problems of residents on the estate: perhaps mothers with young children walking to the centre should not be made to walk further than necessary. The Electricity Board are concerned about transmission losses: what site would minimize these losses? The company is concerned with the cost of moving goods from the factories to the warehouse and from the warehouse to the customers: what site will minimize the total cost of transport?

In each meeting, someone will suggest the 'centre of gravity' as the optimum siting of the buildings. But, how is this to be found, and is it the correct answer? Let us think about how we can represent each of these problems in terms of a mathematical model.

In each case, we measure factors of inconvenience or loss as 'costs' associated with the problem: distance walked for the community centre, power lost in the electricity supply system, money spent on transport in the warehouse problem. The cost will depend on the position of the centre, substation, or warehouse, relative to the positions of houses on the estate or customers of the company. In each case, we can reduce the number of individuals by noting some common paths not dependent on the position of the 'site'. We shall use this word now for each of the three special examples. In the first case many people walking to the community centre will have to walk to some street corner wherever the centre is built. In the second, there will not usually be separate cables from the substation to each house. In the third, goods from the warehouse will pass through the hands of a wholesale merchant. Thus there will be concentrations of demand at these nodes—another suitable word to cover our three cases. In general terms the problem may now be stated thus: given N nodes with n_i individuals concentrated at the ith node, what site will minimize the total cost?

An obvious way of specifying the positions of the site and the nodes is to use geographical map references. Over a sufficiently small part of the area of the Earth, we can regard map references as Cartesian coordinates. Let u, v be the coordinates of the site and x_i, y_i be the coordinates of the ith node at which the concentration is n_i. Let r_i be the distance of the ith node from (u, v) so that

$$r_i^2 = (x_i - u)^2 + (y_i - v)^2.$$

The solution, namely the coordinates (u, v) of the site, will result in a certain cost for each of the n_i individuals at the ith node. As an approximation let us suppose that the cost is the same for each of the n_i individuals. In many practical situations this is quite a reasonable approximation, since the variation in cost between individuals at a node is quite small when compared with variation in cost between nodes. Let the function $g_i(r_i)$ be the cost incurred by an individual at the ith node. Thus the total cost due to the decision is

$$C = n_1 g_1(r_1) + n_2 g_2(r_2) + \ldots + n_N g_N(r_N).$$

We want to find a site (u, v) such that C is minimized. The answer will clearly depend on the form of the functions $g_i(r_i)$. In many cases it is reasonable to suppose that the form of these is the same for all nodes. In practice, the costs will depend on the available road system rather than the straight line distances r_i. Fortunately we can find a practical relationship between road distances and direct distances r if road distances are plotted against direct distances on a graph, it is found that the points lie fairly close to a straight line. There are some obvious exceptions like mountains and journeys through very large cities.

It is interesting to note the solution when all $g_i(r) = kr^2$ with k a constant. This form of $g(r)$ is a reasonable approximation in the case of power losses in an electric cable. To minimize

$$C = \sum_{i=1}^{N} kn_i\{(x_i - u)^2 + (y_i - v)^2\},$$

we differentiate it partially with respect to u and then v (in each case treating all other variables as if they were constants) and set the resulting equations equal to zero:

$$\frac{\partial C}{\partial u} = -2k[n_1(x_1 - u) + n_2(x_2 - u) + \ldots + n_N(x_N - u)] = 0,$$

$$\frac{\partial C}{\partial v} = -2k[n_1(y_1 - v) + n_2(y_2 - v) + \ldots + n_N(y_N - v)] = 0.$$

This is an extension of the simple method for minimizing a function of a single variable. The notation $\partial C/\partial u$, for example, indicates the partial derivative of the function C with respect to u. The best position of the site is therefore

$$u = \sum_{i=1}^{N} n_i x_i \Big/ \sum_{i=1}^{N} n_i, \qquad v = \sum_{i=1}^{N} n_i y_i \Big/ \sum_{i=1}^{N} n_i,$$

the physical centre of gravity of the N weighted nodes. If now we reverse the argument and integrate the differential equations we get back to $g(r) = kr^2$. Thus the best position is at the centre of gravity only when the costs depend on the square of the distances.

What happens if the costs are proportional to distances? Let us take the cost associated with an individual at the ith node to be $s_i r_i$ where s_i is a constant associated with the ith node. Thus C has the form

$$C = n_1 s_1[(x_1 - u)^2 + (y_1 - v)^2]^{\frac{1}{2}} + \ldots + n_N s_N[(x_N - u)^2 + (y_N - v)^2]^{\frac{1}{2}}.$$

This form of C is a reasonable one in the case of the community centre and of the warehouse. The resulting differential equations are much harder to solve.

We are used to the idea of using mathematical equations to predict the behaviour of a physical system. But we can do the reverse, i.e., use a physical system to find the solution of the equations.

Let $n_i s_i = w_i$. Interpret w_i as a weight. Then $w_1 r_1 + w_2 r_2 + \dots w_N r_N$, the sum of the products of weights by distances, can be thought of as the potential energy of a physical system. Since any physical system minimizes its potential energy, the equations can be solved very simply by 'simulation' as follows.

Draw up a map on a table-top and bore holes at the position of each node. Pass strings through the holes with weights hanging down each weight being proportional to w. The strings on the table-top are tied together. The system of strings and weights is then agitated. When it comes to rest, the knot will indicate the best position of the site.

Operational research methods can help us with problems of siting, which might otherwise have been decided by unscientific and possibly unsatisfactory methods. The scope of these methods is immense, some of their most interesting applications being to queueing problems in road traffic and in telephone exchanges.

Whodunit? Or the Reverend Mr Bayes FRS helps to decide

J. GANI
University of Sheffield

During the winter, my children are in the habit of leaving out some slices of bread and bacon for the birds and neighbourhood cats. They know that cats prefer bacon, while birds prefer bread. They have also observed that there are roughly three times as many birds in the neighbourhood as cats. Often, they lay out large supplies of bacon and bread early in the morning, and return in the afternoon to find most of these eaten. My youngest child often asks me: 'Who ate which?' Since I have not been watching all the time, this is a difficult 'Whodunit' to unravel. I shall try to indicate, at least in outline, how this may be done.

To make the problem easier, let us suppose that we are dealing with exactly 3 birds and 1 cat. Having put out exactly 1 piece of bread and 1 slice of bacon, we return to find the bacon eaten, presumably by a single animal. We wish to decide whether it was the cat or a bird. As it stands, the problem cannot be resolved, but if instead of certainty, we are willing to settle for reasonable likelihood, then by recourse to elementary probability theory, we can hope to make some progress with it.

For our purpose, let us regard the probability p of an event A as a positive number within the scale 0 to 1, which expresses the likelihood that A occurs. To say that $p = 1$ is to state that the event A *must* occur, while if $p = 0$, A will not occur. A value $p = 0{\cdot}8$ is equivalent to asserting that the event A has a high likelihood of occurrence, and this could for example be associated with the empirical result that in a thousand trials, say, the event A had occurred 800 times. We might for example decide that the probability of a hungry bird's eating bread is $p = 0{\cdot}8$, while that of its eating bacon is $1-p = 0{\cdot}2$. You will note that we are taking the probability of the bird's eating either bread or bacon as $0{\cdot}8+0{\cdot}2 = 1$, the sum of the probabilities of the two events, which are considered to be mutually exclusive.

We might also decide that the probability of a cat's eating bacon was 0·9, while that of its eating bread was 0·1. We already know that there are 3 birds and 1 cat, and we shall assume that the arrival of any one of these at the feeding trough is equally likely. Denoting by B the event of a bird's arrival, and C the event of the cat's arrival at the trough, the position might be restated in probabilistic terms as

$$\Pr\{B\} = 0{\cdot}75, \quad \Pr\{C\} = 0{\cdot}25,$$

where $\Pr\{B\}$, $\Pr\{C\}$ stand for the probabilities of the events B, C, respectively.

We now require only one further axiom of probability theory. This is one which states that if the event A follows the event B on which it is dependent, then the probability that both events will occur, denoted by BA, is given by

$$\Pr\{BA\} = \Pr\{B\} \times \Pr\{A \mid B\},$$

where $\Pr\{A \mid B\}$ indicates the probability of A given that the event B has already occurred. For example, if A_1 denotes the event of eating bread, and A_2 that of eating bacon, then

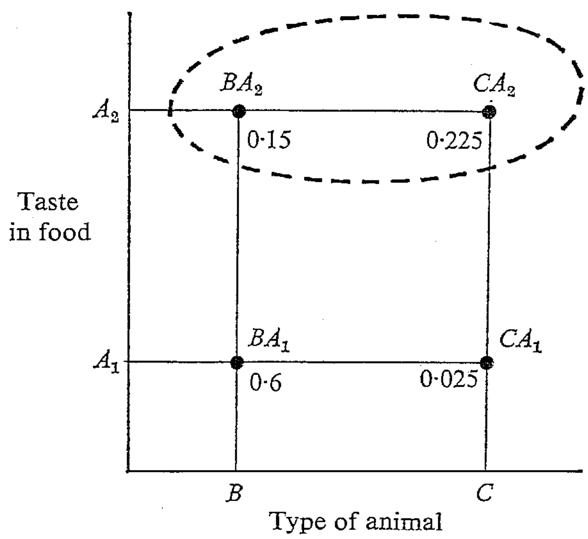
$$\Pr\{CA_1\} = 0{\cdot}25 \times 0{\cdot}1 = 0{\cdot}025,$$

$$\Pr\{CA_2\} = 0{\cdot}25 \times 0{\cdot}9 = 0{\cdot}225,$$

$$\Pr\{BA_1\} = 0{\cdot}75 \times 0{\cdot}8 = 0{\cdot}6,$$

$$\Pr\{BA_2\} = 0{\cdot}75 \times 0{\cdot}2 = 0{\cdot}15.$$

In order to see things more clearly, we may now lay out our set of probabilities in a diagram as follows:



Each of the four heavily dotted points indicates the joint event BA_1, CA_1, BA_2, CA_2, respectively, with its associated probability. Note that the sum of all these probabilities is 1.

Now suppose that we discover after a time that the bacon has been eaten, and we wish to decide which animal has been responsible. We are now in effect stating *after the event* that the probability that the bacon was eaten by cat or bird is 1. But we note that in our diagram the probabilities of the mutually exclusive events BA_2 (bird eats bacon) and CA_2 (cat eats bacon) add up to only

$$\Pr\{A_2\} = \Pr\{BA_2\} + \Pr(CA_2\} = 0{\cdot}375.$$

This probability has, of course, been worked out *before the event*; it is the probability that the bacon is eaten, as against the bread being eaten, when either event may occur. Once *we know* that the bacon has been eaten, then in a certain sense, we remain interested only in the two points indicating the events BA_2, CA_2, and lose interest totally in BA_1, CA_1. Our scale for the probabilities of the events $\{B|A_2\}$, $\{C|A_2\}$, which we have surrounded by an ellipse, is now wrong, since if we know that the bacon has been eaten the probabilities should add up to 1. To correct this, we need to adjust the probabilities by a factor of

$$\frac{1}{\Pr\{A_2\}} = \frac{1}{0{\cdot}375}$$

such that

$$\frac{0{\cdot}15}{0{\cdot}375} + \frac{0{\cdot}225}{0{\cdot}375} = 1.$$

Thus, knowing that the bacon was eaten, we may now obtain that

$$\Pr\{B|A_2\} = \frac{\Pr\{BA_2\}}{\Pr\{A_2\}} = \frac{0{\cdot}15}{0{\cdot}375} = 0{\cdot}4,$$

$$\Pr\{C|A_2\} = \frac{\Pr\{CA_2\}}{\Pr\{A_2\}} = \frac{0{\cdot}225}{0{\cdot}375} = 0{\cdot}6,$$

where these two probabilities add up to 1, indicating the certainty that the bacon was eaten. We note that there is an appreciably greater chance that it was the cat which ate the bacon. I could now answer my child: 'It looks as if the cat ate the bacon rather than the bird.'

What we have presented in a very simple form, is a theorem on inference due to the Reverend Mr Thomas Bayes FRS, an English mathematician of the eighteenth century (1702–61). His theorem is concerned with finding the probability of an event, after a particular outcome in a trial, given prior probabilities of events before the trial. We could state Bayes' theorem mathematically as follows:

$$\Pr\{C|A_2\} = \frac{\Pr\{C\}\Pr\{A_2|C\}}{\Pr\{B\}\Pr\{A_2|B\} + \Pr\{C\}\Pr\{A_2|C\}}.$$

All the probabilities on the right-hand side are known before the trial, while that on the left obtained from them is dependent on the outcome of the trial. A formal demonstration of this theorem can readily be given. We need only accept that the events CA_2 and A_2C are identical, so that

$$\Pr\{CA_2\} = \Pr\{A_2C\}.$$

Thus, formally

$$\Pr\{A_2C\} = \Pr\{A_2\}\Pr\{C|A_2\},$$

so that

$$\begin{aligned}\Pr\{C|A_2\} &= \frac{\Pr\{A_2C\}}{\Pr\{A_2\}}\\ &= \frac{\Pr\{CA_2\}}{\Pr\{BA_2\}+\Pr\{CA_2\}}\\ &= \frac{\Pr\{C\}\Pr\{A_2|C\}}{\Pr\{B\}\Pr\{A_2|B\}+\Pr\{C\}\Pr\{A_2|C\}}.\end{aligned}$$

This gives us our result. Some people find it easier to draw diagrams and reason from them than to write out the formula in full, but both methods lead to the same conclusion.

It may be of some interest to learn that Thomas Bayes, born into a Nonconformist family in 1702, was ordained into the Presbyterian ministry in his early twenties. While serving as a minister at Tunbridge Wells, he took a serious interest in mathematics and was elected to the Royal Society in 1742. He published little, but what he wrote was of the very highest quality. Concerning his article which contained what is now referred to as 'Bayes' theorem', he had very serious doubts, and refused to make his results public. It was found among his papers after his death in 1761 by a friend, Richard Price, who had it published posthumously in 1763 under the title 'An essay towards solving a problem in the doctrine of chances', in the *Philosophical Transactions of the Royal Society* **53** (1763) 370–418. Those who are interested in this paper will find a modernized version of it, together with a biographical note on Thomas Bayes by Professor G. A. Barnard of the University of Essex, in *Biometrika* **45** (1958) 293–315.

Bayes' theorem is of fundamental importance in the theory of statistical inference. An example of its practical use may help to illustrate this. Two telephones B and C on the same party line are in use 0·05 and 0·3 of the total available time respectively. B is old and breaks down with probability 0·2 per unit time in use, while C is new and breaks down only with probability 0·04. A failure indicator (F) shows that one of the two phones is out of order: which is it more likely to be?

Bayes' theorem gives us that

$$\Pr\{B|F\} = \frac{\Pr\{BF\}}{\Pr\{BF\}+\Pr\{CF\}} = \frac{0{\cdot}01}{0{\cdot}01+0{\cdot}012} = \frac{5}{11},$$

$$\Pr\{C|F\} = \frac{\Pr\{CF\}}{\Pr\{BF\}+\Pr\{CF\}} = \frac{0{\cdot}012}{0{\cdot}01+0{\cdot}012} = \frac{6}{11}.$$

Thus, we find that the newer telephone is slightly more likely to break down; this would be of value in deciding where an engineer to repair it should call first.

For those who would like to try their hand at a harder party trick, the following problem may be of interest.

Three boxes A, B, and C contain

A: 4 red balls,

B: 1 white and 3 red balls,

C: 2 white and 2 red balls.

You are blindfolded, and the boxes are shuffled around until one is put on your right hand, one in the centre, and the last on your left. You are now asked to pick one ball from whichever box you prefer. Suppose you do this and you are told that the ball you pulled out was red, can you decide which of boxes A, B, or C the ball is likely to have come from? (The answer will be found in the next issue.)

Mathematics and the Physical World

G. T. KNEEBONE
Bedford College, University of London

Mathematics, as we first meet it, is a kind of skill, and the test of our mastery of its methods is whether we are able to solve problems. But by degrees the centre of our interest shifts, and problems give way to theorems. The theorems turn out to be what we are chiefly interested in, and problems are seen to be valuable only as an aid to fuller understanding of theorems. Mathematics thus comes to be treated as a body of factual information—about numbers, geometrical figures, sets, or mathematical objects of some other kind; but the 'facts' of mathematics are set apart from the facts of everyday life or scientific theory by their timelessness and complete certainty. As man's knowledge of the world grows, most other factual knowledge becomes obsolete, and has to be discarded or revised. Thus the Earth was once held to be the centre of the universe, and at a later time all matter was thought to be made up of indivisible atoms, but both these beliefs have now been given up as erroneous. The mathematical assertion that $2+2=4$, on the contrary, is no less true today than it was in the earliest times, and that it should ever be disproved is inconceivable. What, then, is the nature of the 'facts' of mathematics, and why are these exempt from change?

In order to discuss this question we have no need to go into higher mathematics, since even simple arithmetic is already wholly typical. Let us then confine the argument to the natural numbers (that is to say, the numbers 1, 2, 3,... that are used in counting) and the two operations of addition and multiplication that may be carried out on these numbers. Our earliest knowledge of the natural numbers and of how to manipulate them is gained in a very elementary way—for instance by playing with heaps of pebbles—and at that stage the facts of arithmetic are simply items of everyday knowledge, no more exalted than the rest of what we have discovered about the world in which we find ourselves. But arithmetic soon becomes a matter of calculation rather than of direct observation, and an element of generality thus creeps in. When this happens, arithmetical statements begin to acquire certainty of a rather special kind.

Although reliance on calculation brings to arithmetic a certain detachment from empirical observation, we find that when arithmetical conclusions are reached in this way their practical value is enhanced rather than diminished; and abstract arithmetic continues to have direct application to the actual world. This is what is puzzling. How arithmetical truths can be known with complete certainty, and yet at the same time give information about the empirical world, is a famous and much-discussed philosophical problem; but fortunately a first approximation to a solution, which may help to make the nature of mathematics a little clearer, is possible in quite simple terms. Briefly, then, a fact can only be arrived at by pure deduction if it is a consequence of the definitions of the words used in expressing it; and such is indeed the case with the facts of arithmetic. The number of technical terms ordinarily used in arithmetic is not at all large, and it becomes even smaller when a few of these terms are singled out as basic and the others are defined by reference to them. If we take '$+$' as basic, for instance, we can define '$<$' by expressing '$a < b$' in the alternative form 'there is an x such that $a+x = b$'. It turns out that, for quite a substantial part of the arithmetic of the natural numbers, we can define everything that we need in terms of only five basic notions: number, the particular number 1, the successor of a number (i.e. the next number in order of magnitude), and the operations symbolized by '$+$' and '$\times$'.

The one essential feature of the natural numbers, from which all else follows, is that these numbers make up an unlimited sequence or progression. More precisely, there is a starting number 1, and the entire totality of numbers can be generated from this one by repeated passage from the number n last obtained to its successor n'. We are thus led to take as basic (i) the concept of natural number itself, and (ii) the specific number 1 and the operation which leads from n to n'. These concepts can be defined implicitly by five declarations, usually referred to as the *Peano axioms*,† which together embody all that we shall presuppose in proving arithmetical theorems. The Peano axioms may be formulated as follows:

(1) 1 is a natural number.
(2) If n is a natural number, then so also is its successor n'.
(3) 1 is not the successor of any natural number.
(4) If $a' = b'$, then $a = b$; i.e., no two numbers have the same successor.
(5) If a property P is such that (i) it belongs to the number 1, and (ii) whenever it belongs to a number n it belongs also to n', then P belongs to every natural number.

† After Giuseppe Peano (1858–1932), an important Italian mathematician and logician. For further information, see G. T. Kneebone, *Mathematical Logic and the Foundations of Mathematics* (Van Nostrand, 1963), Chapter 5.

It follows from (1)–(4) that there is an unending succession of ever-different natural numbers, and from (5) that every natural number occurs somewhere in it.

Axiom (5) is known as the *axiom of induction*, since it yields the fundamental method of proof by mathematical induction. Suppose we have a mathematical statement, conveniently symbolized by $\Phi(n)$, which makes an assertion about an unspecified natural number n, for instance the statement $1+2+\ldots+n = \frac{1}{2}n(n+1)$. To prove by induction that $\Phi(n)$ is true for every n, we show two things: (i) $\Phi(1)$ is true, and (ii) if $\Phi(n)$ is true, then $\Phi(n')$ is true also. This is sufficient to establish the conclusion that $\Phi(n)$ is true for all n. For let P be the property which a number n possesses if and only if $\Phi(n)$ is true. Then, by (i) and (ii), P is a property of 1 and it is a property of n' whenever it is a property of n. It follows from the induction axiom that P is a property of every natural number, that is to say that $\Phi(n)$ is true for all n.

To define the operations $+$ and $\times$, we also proceed in an inductive manner, taking as further axioms the two pairs of equations

$$a+1 = a', \qquad a+n' = (a+n)' \quad \text{and} \quad a\times 1 = a, \qquad a\times n' = (a\times n)+a.$$

These *equations of recursion*, as they are called, do indeed determine $a+b$ and $a\times b$ unambiguously for any particular a and b. Thus, for example, since $2 = 1'$, $3 = (1')' = 1''$, and so forth,

$$3+2 = 1''+1' = (1''+1)' = (1''')' = 1'''' = 5,$$

$$2\times 2 = 1'\times 1' = (1'\times 1)+1' = 1'+1' = (1'+1)' = (1'')' = 1''' = 4.$$

When arithmetic is made to depend in this way on a few clearly stated first principles, and when we mean by an arithmetical theorem simply an assertion which follows logically from these principles by a chain of deductive argument, it is evident that all the theorems will in a sense be eternal truths. But this is so only within the abstract world of thought defined by the axioms, a 'world' that must be clearly distinguished from the concrete world of experience, to which abstract arithmetic may possibly be applied. To justify such an application we need to produce suitable concepts with empirical reference, and to show that if the technical terms of arithmetic are taken as referring to these concepts, then the arithmetical axioms are all true assertions about the actual world. Now the familiar intuitive conception of number as a measure of the size of a finite collection of things makes sense for a world in which we are able to distinguish separate objects which retain their individuality. If we understand 'number' in this sense, if we identify '1' with the measure of any collection which has something belonging to it but does not have any distinguishable things belonging to it, if by the successor of a number n we mean the measure of a collection formed by the addition of a further member to a collection of measure n, and if we take the operations symbolized by '$+$' and '$\times$' to have their normal intuitive meanings, the axioms of arithmetic are then true statements about the world. It follows that all conclusions that are correctly deduced from the axioms are also true statements about the world; and thus abstract arithmetic, when suitably interpreted, yields empirical truth.

Professor Pólya on Intuition and Learning

George Pólya, Emeritus Professor at Stanford University, has always emphasized both in his teaching and in his books the importance of intuition in mathematical discovery. The following three aphorisms selected by him illustrate his views on intuition.

> Intuition is the conception of an attentive mind, so clear, so distinct, and so effortless that we cannot doubt what we have so conceived.
>
> DESCARTES.

> Nothing is more important than to see the sources of invention which are, in my opinion, more interesting than the inventions themselves.
>
> LEIBNITZ.

> Thus all human cognition begins with intuitions, proceeds from thence to conceptions, and ends with ideas.
>
> I. KANT.

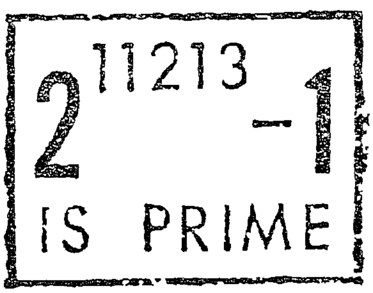
For a detailed account of Pólya's ideas, readers may consult his two-volume work *Mathematical Discovery* (John Wiley & Sons, New York, 1962).

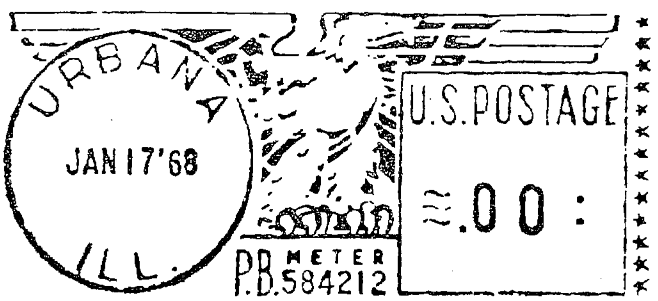
Computers and their Uses

J. HOWLETT

Atlas Computer Laboratory, Chilton, Berkshire

In 1963, the Mathematics Department of the University of Illinois had a stamp made with which it could put on its mail the message '$2^{11213}-1$ is prime'. This is the largest known prime number; it has 3376 digits, beginning 281411... and the fact





that it is a prime was established with the aid of the university's electronic digital computer. Finding this out obviously involved a very great deal of work, but is something which, in a general way, we can understand being done by a machine. This is just arithmetic, computers do arithmetic, so we ought to be able to make a computer do this for us. In the same way, working out π to a few thousand places—by summing various series derived from $\tan^{-1} x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 \ldots$—is impressive but quite comprehensible. But computers are now practically a household word and we are apt to see in the daily press statements like 'Computer Predicts Weather' or 'BR Time-tables now produced by Computer' or 'Computer Questions Authorship'. How can this kind of thing be done? And can we say what can be done with a computer and what cannot?

To answer such questions we need to take a very basic look at the computer and what it does. It is rather a pity that the name computer is used—it is here to stay of course—because it suggests that it just does arithmetic: 'information processing machine' is a better name and for a start we can think of it as a black box which can receive information from the outside world, perform some operation on it, and put out a response. The information can be collections of digits, letters, and symbols arranged so as to form numbers, words, and statements. For example, one item of information could be a pair of numbers and the instruction to add them together and print the sum. Or it could be a list of values of the coefficients (a_{ij}) and right-hand sides (b_i) of a set of linear equations

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i = 1, 2, \ldots, n),$$

together with the rules of solving for the (x_i) and the instruction to print the solution. Or, again, to arrange a group of words into alphabetical order. Being a machine it has to receive these instructions in some form of code which will cause its (electronic) components to act so as to carry out the instructions. There is nothing unfamiliar in this: the telephone codes the digits which we dial into trains of pulses which cause relays to open or close so as to connect our telephone with that of the person to whom we wish to speak, or to give us in code the information that someone else is speaking to him already.

To be just a little more detailed, the essential parts of a computer are the *store*, in which information is held, and the *processor*, which performs the operations on this information. The *input* and *output* units link these with the outside world. Two very basic facts underlie the whole philosophy of the computer. One is that any piece of information can be represented (i.e., coded) by a sequence of binary digits—a sequence consisting of 1's and 0's. The other is that a binary digit itself —usually called a *bit*—is very easily represented electronically, for example by a train of pulses each of which can be positive or negative (positive representing 1 and negative 0, say), or the kind of circuit called a flip-flop which has two states of equilibrium, or by a magnetic element which can have its polarity in either of two directions. In modern machines the store is made up of magnetic elements, usually tiny rings of a ferrite material, called *cores*, and the processor of transistors, diodes, capacitors and resistors. The machine gets its generality, that is, the ability to tackle many different types of problem, from the property of representing information in binary code, and its power from the great speed with which electronic events take place, up to many millions a second.

Of course the user of a computer does not need to think about all this electronics and binary coding when he wants to solve a problem, any more than when he dials a telephone number. It was otherwise in the early days but now computers can be made to operate in response to formalized but fairly natural-looking instructions. A large number of languages have been devised for communicating with computers, of which the two most widely used throughout the world are called Fortran (short for Formula Translation) and Algol (Algorithmic Language). Suppose we want to make the machine find the greatest of a string of numbers—a simple task, but not trivial. Call these $x_1, x_2, \ldots, x_n$ and the maximum *x*max. Then in the language Algol this program of instructions will cause the machine to pick out *x*max:

```
real procedure xmax(x, n)
array x; integer n;
begin
    integer i,j; j: = 1
            for i: = 2 set 1 until n do
            if x[i] > x[j] then j: = i;
            xmax: = x[j];
end.
```

The words 'real procedure', 'array', 'integer' give information about the kinds of quantity we are working with. The instruction

for i: = 2 step 1 until n do

means 'obey all the following instructions in which i appears, giving it the values 2, 3,..., n in turn'.

We type this program on a machine which has a keyboard like that of a typewriter but simultaneously produces a pack of punched cards or reel of punched paper tape, on which the characters—letters, digits, punctuation and other symbols—are coded into patterns of holes. When the cards or tape are placed in the computer's input mechanism, photo-electric cells detect the position of the holes and so translate them into patterns of pulses, and the machine thus gets the binary representation of the information which it needs in order to solve the problem. To complete the job we would have to give it also, on cards or tape, the list of numbers (x_i)—which could be of any length—and the number n of these, and instructions to print out the answer.

This is very simple, but illustrates the most important points about the use of the computer. However long and complicated the problem we want to solve, we must know exactly what it is we want to do, and be able to write down a sequence of absolutely precise instructions which will lead to the answer. The example uses only one kind of operation on the numbers $x_1, x_2, \ldots$, that of comparison; computers are built so that they will do all the simple arithmetical operations, and from these we can construct more complicated processes: thus we could find the sum of the squares of the numbers, and then the square root of this, or the sine of the largest or smallest of the numbers. They are built also to do operations which are not arithmetical, such as comparing the individual binary digits representing two characters or combining these in various ways to produce a third. These enable us to operate on characters and symbols in ways which have nothing to do with arithmetic: comparing two words to see if they are identical, or sorting a string of words into dictionary order, or doing operations of ordinary algebra like simplifying expressions.

We can now try to answer the questions I raised in the first paragraph. The short answer is that if we know what it is we want and can describe a process which will give it us, we can write a computer program to carry out the process and give us the answer. If it is a question of evaluating expressions, however complicated, there is no real problem: working out π from the series is an example; the same goes for solving linear equations—there are perfectly definite processes such as successive elimination and back-substitution which we can program. Many problems in physics and engineering face us with finding the numerical value of a definite integral or solution of a differential equation, and these are more difficult. We can seldom find an exact algebraic solution into which we can substitute numbers, but we can use the methods of a branch of mathematics called the Calculus of Finite Differences, especially developed to attack problems of this kind. These give us processes which we can program, but they can raise other problems: they involve us in making approximations and the effect of these on the solution can be exceedingly difficult to assess. With so much use being made of computers, this is a live subject for current research; far from reducing the need for a good understanding of mathematics, the computer has made this greater than ever. The problems of weather-prediction and railway time-tabling are of this kind. Weather depends on the motion of the atmosphere, which can be described by a set of partial differential equations, so that forecasting is really a problem of integrating differential equations. Unfortunately the equations are so complicated that even with the fastest computers yet produced we cannot get the solution quickly enough to keep up with the weather; moreover, there is no finality about the equations themselves because meteorological processes are still not fully understood. But computers are being used in England and America to help with weather prediction and will certainly be used more and more in the future. The railway-train problem is much simpler. The motion of the train along the track satisfies a differential equation; given enough information about gradient, curvature, speed restrictions, air-resistance, and power/speed characteristics of the locomotive we can program the computer to produce a numerical solution, which is a set of point-to-point timings.

The remaining example—computers and authorship, about which quite a lot has been written—takes us away from the quantitative problems of science and into a field where we are not so sure of what we want. The computer, as I said earlier, can do non-numerical operations just as readily as numerical. In the example, it picked out the largest number from a list; if the list had been one of words instead of numbers we could have written a program to pick out one according to any precise criterion—with the smallest number of letters, or the first in dictionary order or the same as another given word. This simple fact is the clue to the use of the machine to select information from any kind of list. If each item of the array—corresponding to the (x_i)—is a group of words, say a list of book titles with the author's name first, we can make it pick out all the books by any given author; and with a little more trouble we can pick out all titles which contain any given word—all titles mentioning computers, for example. We actually do this in the Atlas Laboratory, where we keep the library catalogue on the computer; we also have an index of all the papers in the leading journals of computer science and can use the machine to find all those by any named author, together with references to all the papers which he quotes and all the later papers which refer to his work—this is a great help in searching the literature for some piece of information. This, however, is all perfectly definite and therefore programmeable; if we had said we wanted a list of the 'best' books, or the 'most poetic' words we would not have known how to go about it, because we have no objective criteria for either of these concepts. The computer has nothing to offer us now. The authorship problem is not as extreme but certainly has some of this difficulty. The problem is this: authors have styles—as do artists and musical composers—which we as humans can recognize and use to identify the origin of a piece of written work. Can we find any quantitative measures which we can use to give an objective specification of an author's style, and to settle questions of disputed authorship? We can make the computer scan pieces of text and give us any objective information we care to ask for: total number of different words (i.e., total vocabulary), frequency of occurrence of any particular word or group of words ('rosy-fingered dawn' of Homer, for example) or kind of word, such as prepositions, or distribution of sentence length, or anything else. Literary scholars are now finding it possible, thanks to the computer, to get information of this kind which previously would have meant years of drudgery. But style is a very subtle concept and I personally feel that the interpretation of this information is very much the job of the scholar, with the computer simply providing the numerical facts.

Finally, a few words about speed. The first electronic computers were being manufactured in the early 1950's—the very first to be sold was the Mark 1 made by Ferranti Limited for Manchester University in 1951—and would do one instruction in a few milliseconds: the Mark 1, for example, would do a multiplication in 2 milliseconds. Overall, they would do complete calculations about 1000 times faster than anyone could do by hand with an ordinary desk calculator. This was already a startling increase in speed—the motor-car is only about 10 times the speed of the horse, but has revolutionized our lives; and the jet plane is only another factor of 10 beyond that, or 100 altogether. Modern machines have instruction times of a few microseconds, and are at least 1,000,000 faster than hand calculation, a factor which is very hard to appreciate. With our machine at Chilton we calculated, as a test, the value of π to 5000 decimals in 20 minutes, and $\sqrt{2}$ to 10,000 decimals in 14 minutes: either of these would have needed a life-time of hand calculation. We can solve a set of 100 simultaneous linear equations in 100 unknowns in 15 seconds; to do this by hand would take about 2 years. All this means that we can now embark on calculations which were unthinkable before the computer came along. It also means that we are only at the beginning of the computer age and have a great deal to learn about the powers of this remarkable invention.

From Rule of Thumb to Abstract Structure

L. MIRSKY
University of Sheffield

'Every schoolboy knows who imprisoned Montezuma and who strangled Atahualpa': it may be doubted whether Lord Macaulay's faith in the erudition of the British schoolboy was entirely well-founded. However, if for the faintly esoteric item of historical information, he had substituted the claim that 'every schoolboy knows how to test a number for divisibility by 3', one could hardly question his dictum. The test is, indeed, known to almost everyone; but behind this fact there lurks a further query: in any group of 100 people familiar with the test, how many could offer reasonable grounds for accepting its validity? A likely guess would be 'very few'; and yet the underlying mathematical ideas are thoroughly elementary, very easy to grasp, and at the same time important since they lead not only to an understanding of the properties of the number system but also serve to illustrate algebraic concepts such as those of group, ring, ideal, and integral domain which pervade a very large part of modern mathematics.

The test for divisibility we have mentioned states that *an integer is divisible by 3 if and only if the sum of its digits* (*when it is written in the scale of* 10) *is divisible by 3.* To prove this result, we begin by recalling that, after dividing an integer a by a positive integer m, we are left with a (uniquely determined) remainder which is equal to one of the numbers $0, 1, \ldots, m-1$. If two integers a and b leave the same remainder when divided by m, we say that *a is congruent to b modulo m* and we write the relation symbolically as $a \equiv b \pmod{m}$. In particular, then, the relation $a \equiv 0 \pmod{m}$ means that a leaves the same remainder as 0, i.e. a leaves the remainder 0. In other words, $a \equiv 0 \pmod{m}$ means that a is divisible by m.

A relation of the form $a \equiv b \pmod{m}$ is called a *congruence*, and it is very easy to develop a simple algebra of congruences. In the first place, we have the following result.

THEOREM 1. *If $a \equiv a' \pmod{m}$ and $b \equiv b' \pmod{m}$, then*

$$a+b \equiv a'+b' \pmod{m}, \qquad ab \equiv a'b' \pmod{m}.$$

The proof is easy and we leave it as an exercise for the reader.

Now let N be a given (positive) integer, and let it be written in the scale of 10, so that

$$N = a_0 + 10a_1 + 10^2a_2 + \dots + 10^k a_k,$$

where each of the digits $a_0, a_1, \dots, a_k$ is equal to one or other of the numbers $0, 1, 2, \dots, 9$. Since $10 \equiv 1 \pmod 3$, we infer by repeated use of the multiplication formula in Theorem 1 that $10^2 \equiv 10 \equiv 1$, $10^3 \equiv 10 \equiv 1, \dots \pmod 3$. Hence, again by this formula,

$$a_0 \equiv a_0, \quad 10a_1 \equiv a_1, \quad 10^2a_2 \equiv a_2, \quad \dots, \quad 10^k a_k \equiv a_k \pmod 3$$

and therefore, by the addition formula in Theorem 1,

$$N = a_0 + 10a_1 + 10^2a_2 + \dots + 10^k a_k \equiv a_0 + a_1 + a_2 + \dots + a_k \pmod 3.$$

Write $d = a_0 + a_1 + \dots + a_k$, so that d is the sum of the digits of N. We have then shown that $N \equiv d \pmod 3$, i.e. N and d leave the same remainder when divided by 3; in particular, N is divisible by 3 if and only if d is divisible by 3. This establishes the rule quoted a little earlier.

There is, as is generally known, an analogous rule for divisibility by 9: its demonstration presents no novel features. A test for divisibility by 11 requires a slightly more sophisticated formulation, but the proof can be based on the same ideas as those used above. The reader may like to attempt to state and prove this test.

Tests for divisibility, though amusing, are of course only of limited interest; but the notions introduced for dealing with them are fundamental. Now one of the most obvious facts about the development of mathematics is that invariably an initial analysis of simple objects leads step by step to the construction and manipulation of more complex objects built out of simple ones. This theme is readily illustrated from the theory of congruences. In this case, starting with integers, we ultimately learn how to handle not only these but whole classes of integers, namely the *residue classes*. A residue class modulo m is the class of all integers which leave the same remainder when divided by m. For example, when $m = 5$, there are 5 residue classes, and one of them consists of the integers

$$\dots -13, -8, -3, 2, 7, 12, \dots.$$

In what follows, we shall take m to be a fixed positive integer. If a is any integer, we shall denote by $[a]$ the residue class consisting of all those (infinitely many) integers which leave the same remainder as a on division by m. We note that $[a] = [a']$ precisely when a, a' leave the same remainder, i.e. $a \equiv a' \pmod m$. Let A be a residue class. Any integer a such that $A = [a]$ (i.e. any integer a which belongs to A) is called a *representative* of A.

We have now constructed m new objects, namely the m residue classes modulo m:

$$[0], [1], [2], \dots, [m-1]. \tag{1}$$

Our next task is to imprint some structure upon this collection by defining certain operations on the residue classes. The operations in question—there are two of them—will be called 'addition' and 'multiplication' since they are closely related to addition and multiplication as understood in the ordinary arithmetic of integers.

Let, then, A and B be residue classes modulo m. We now wish to define two new classes, which we shall denote by $A+B$ and AB (and which are called the sum and product respectively of A and B). Very natural definitions suggest themselves almost at once. Suppose that a is a representative of A and b a representative of B, so that $A = [a]$, $B = [b]$. We now define $A+B$ as the residue class of which $a+b$ is a representative. In other words, $A+B = [a+b]$ or, equivalently,

$$[a]+[b] = [a+b]. \tag{2}$$

Multiplication of residue classes is defined analogously by the formula

$$[a]\,[b] = [ab]. \tag{3}$$

But are these definitions of sum and product really free from ambiguity? After all, both A and B have many representatives, and if for example definition (2) is to make sense, we must have $[a+b] = [a'+b']$ for any representatives a', b' of A, B respectively. Happily, this is the case. For $[a] = [a']$, $[b] = [b']$, and this means that $a \equiv a'$, $b \equiv b' \pmod{m}$. Therefore, by the addition formula in Theorem 1, $a+b \equiv a'+b' \pmod{m}$, i.e. $[a+b] = [a'+b']$, as required. Similarly, we can verify the consistency of the definition (3) by invoking the multiplication formula in Theorem 1.

Having defined addition and multiplication of residue classes, we proceed to examine the properties of these operations. It is not possible to undertake here a systematic study, and we shall confine ourselves to a single question: do the classes (1) constitute a group? Put in this form the question lacks precision since 'group multiplication' has not been specified. We therefore ask, in the first place, whether the classes (1) constitute a group under the *addition* of residue classes. It is readily seen that this is, indeed, the case. The unit element of the group is [0] and the inverse of $[a]$ is $[-a]$. We omit the rather uninteresting details and turn to a much more pointed question: do the classes in (1) constitute a group under *multiplication* of residue classes? The answer to the question as it stands is certainly 'No', for we have $[1][0] = [0][0]$ ($= [0]$) and if the classes formed a group it would follow that $[1] = [0]$, i.e. $1 \equiv 0 \pmod{m}$, which is false (except when $m = 1$). The villain here is the class [0] and we therefore make a fresh start by excluding it and asking whether the classes modulo m:

$$[1], [2], \ldots, [m-1] \tag{4}$$

form a group under multiplication. The answer is 'Not necessarily'. Thus, when $m = 6$, we have $[2][3] = [0]$, so that multiplication takes us outside the collection (4).

A further restriction is therefore necessary, and we shall postulate that m is a prime (i.e. a number greater than 1 whose only divisors are 1 and m). Now let $[a]$, $[b]$ be classes among those enumerated in (4). Then $[a][b]$ is again found in (4). For assume, on the contrary, that $[a][b] = [0]$, i.e. $[ab] = [0]$. This means that $ab \equiv 0 \pmod{m}$, i.e. m divides ab. But m is a prime and it is one of the basic facts of arithmetic that, if m divides ab, then it must divide at least one of the numbers a, b. Thus $[a] = [0]$ or $[b] = [0]$, and this is certainly false since both $[a]$ and $[b]$

are taken from (4). We have therefore demonstrated that the collection (4) is closed under multiplication.

Again, the associative law

$$([a][b])[c] = [a]([b][c])$$

is evidently valid as each side is equal to $[abc]$. Further, the class [1] plays the part of the unit (or neutral) element since, for each class $[a]$ in (4),

$$[a][1] = [1][a] = [a].$$

It remains to establish the existence of inverses. To do this, we need to quote the following simple result.

THEOREM 2. *Let a be an integer and m a positive integer. Then there exists an integer x such that $ax \equiv 1 \pmod{m}$ if and only if the highest common factor of a and m is equal to* 1.

We omit the proof which is not difficult. It may be found in many books; see, for example, page 54 of *Theory and Problems of Modern Algebra* by F. Ayres, Jr. (Schaum Publishing Company, New York, 1965).

Now, if m is a prime and a is any one of the numbers $1, 2, \ldots, m-1$, then the highest common factor of a and m is 1 and so, by Theorem 2, there exists an integer x such that $ax \equiv 1 \pmod{m}$, i.e. $[ax] = [1]$ or, equivalently,

$$[a][x] = [x][a] = [1].$$

Now clearly $[x]$ is different from [0] and so $[a]$ has an inverse element $[x]$ among the classes in (4). We have thus verified that, when m is a prime, (4) is a group under multiplication of residue classes.

This is not the end but the beginning of the road; and the reader who wishes to continue the journey has the choice of many guide books. One of these, in which the structure of the number system is discussed in the wider context of general algebraic ideas, is the book by Ayres referred to a few lines earlier.

Space

M. J. LIGHTHILL
Imperial College, University of London

Although this country is not involved in the disproportionately costly publicity-seeking business of putting human beings in space, there are good opportunities in Britain in the field of space science and technology, involving the far less costly use of instruments of many kinds in spacecraft for practical and scientific purposes. The space engineering work is run by the Ministry of Technology and the space science work by the Science Research Council.

Why is the word 'space' used? There are two good reasons. First, it reflects a feeling of emptiness, correctly because the air pressure halves every 6 km or so as you go upwards, and is down to a millionth of its sea-level value (a very low pressure) at an altitude of about 120 km (1 per cent of the Earth's diameter). The atmosphere, then, is a very thin covering of air, as thin as an apple-skin is in relation to the diameter of the apple.

Secondly, 'space' conveys a sense of three-dimensionality. Human beings living in the atmosphere (or flying within it in aircraft) inhabit a very thin, two-dimensional apple-skin surface. There is no part of it from which you can see much of the rest of the skin; any straight-line path between widely separated points is obstructed by the solid interior. A spacecraft, however, can get right away from the surface and look down on it in 3D; furthermore, communication is possible along dog-leg straight-line paths that join points on the Earth's surface via some spacecraft.

Both reasons for the name 'space' suggest good reasons for sending objects up there. The high vacuum is ideal for astronomical research. The Sun and the stars emit all sorts of informative radiation (ultra-violet, infra-red, X-rays, and so on), in addition to those few sorts (mainly visible light and radio waves) which the atmospheric skin round the Earth allows to reach the surface. At last telescopes in spacecraft are picking up those other radiations and sending pictures of them back to Earth, leading to new knowledge of the universe as revolutionary as, ten to twenty years ago, the first radio-telescopes uncovered.

More important from a practical standpoint is the 3D view of the Earth's surface. Meteorological satellites are now taking marvellous pictures of cloud cover over different parts of the world, and transmitting them on a prompt and regular basis to the local weather-predicting services, making possible big improvements in forecasting. Navigators who traditionally fixed their position by the stars can now do so much more reliably by observation of special satellites, that make radio transmissions so that they can still be 'seen' in cloudy weather.

Early long-distance communication sent messages along wires in a tap-tap code, conveying information by the presence or absence of an electric pulse within each of a regular series of time intervals. Modern communication uses radio waves, in which there is a vastly faster 'tap-tap' as the electric field in a certain direction switches on and off thousands (or millions) of times a second. The wave is transmitted in a direction at right angles, and the frequency of taps still determines how much information per second can be conveyed.

Powerful radiation from the Sun, and also the streams of charged particles which it emits (whose trapping by the Earth's magnetic field has been uncovered by specially instrumented spacecraft) generate a complicated layer of ions and electrons at the top of the Earth's atmosphere. Only this layer, the ionosphere, makes ordinary long-distance radio communication possible. Electrons in it, being exceedingly light, dance to and fro as the tap-tap of the electric field comes on and off, and this dance causes the radio waves to bounce back, and so curve round inside the apple-skin by multiple bounces between ionosphere and Earth's surface.

But electrons, incredibly light though they are (9×10^{-31} kg), just cannot get moving when the frequency of taps is more than a few million times a second. Modern technological requirements, and the complex organization of the modern world, demand that information be conveyed over long distances at rates greater than that, but the appropriate higher-frequency waves shoot through the ionosphere on perfectly straight paths. However, if a communications satellite is up there, it can absorb such radio waves, transmitted along a straight path with frequencies of *thousands* of millions a second, and re-transmit them to a receiver on quite a different part of the Earth's surface, with enormous gains in the amount of information transmitted.

Spacecraft can also explore, and unmanned ones have proved very effective for this purpose. What objects are there to investigate (besides the particles and radiation and magnetic fields that have already been referred to)? Only those objects which, like the Earth, go round the Sun; that is, the eight planets listed in

TABLE 1. The Earth and the eight planets, listed in order of the radius of the approximate circle which each describes (all in one plane) around the Sun

Name of planet	Distance from Sun (km)	Approximate diameter (km)	Number of moons going round it
Mercury	58,000,000	5,000	0
Venus	110,000,000	12,000	0
Earth	150,000,000	12,700	1
Mars	230,000,000	6,800	2
Jupiter	780,000,000	140,000	11
Saturn	1,400,000,000	115,000	9†
Uranus	2,900,000,000	51,000	4
Neptune	4,500,000,000	50,000	1
Pluto	6,000,000,000	13,000	0

For comparison, the Sun has diameter 1,400,000 km and our Moon diameter 3,500 km (and distance from the Earth 380,000 km).

† Together with the famous "rings".

Table 1, the moons which in turn go round them (we have just one, but Jupiter has eleven, of which Galileo discovered four already when he first had the idea of turning the recently invented telescope on to the heavens), and various much smaller objects (asteroids, comets, and meteors).

Even Pluto, at its vast distance of 6,000,000,000 km, might conceivably be reached by spacecraft one day; this is quite impossible, however, for any of the incredibly numerous array of stars and galaxies, not in motion round the Sun, which are so prominent in the heavens. The solar system (of which Table 1 lists all the principal occupants) is, in general, explorable. Objects outside it, of which the nearest (Proxima Centauri) is 30,000,000,000,000 km away, are for various reasons completely out of possible reach. 'Space', then, for practical purposes *is* the solar system.

In that system, why are there all these objects *going round* other objects? Such rotation without physical connection is quite outside our everyday experience. You can whirl an object round on a string, but only the tension in the string allows the circular motion to be maintained. This is because, in a time t, the direction of motion turns through an angle (see Figure 1) producing a change in momentum (the mass m of the object times its directed velocity v) from mv in one direction to mv at this angle to it.

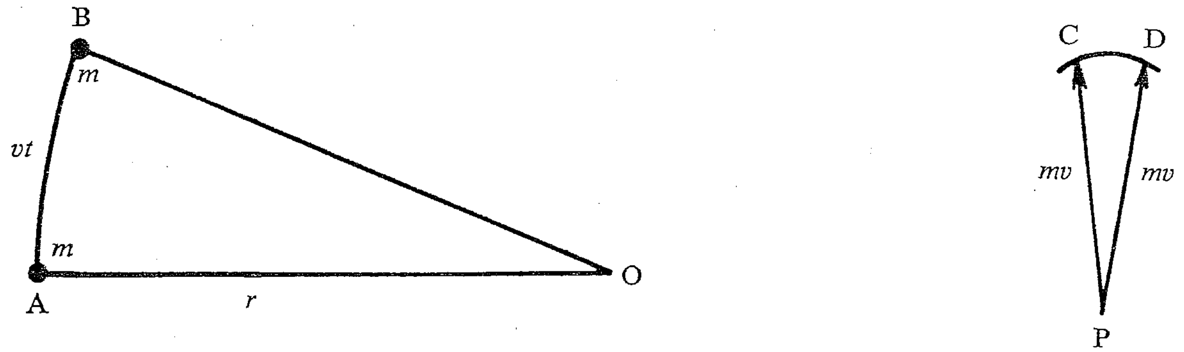


Figure 1. Whirling of mass m about O on a string OA, that after time t is in the position OB. Left-hand arrow PC represent momentum of mass at A, while PD represents momentum after time t when it is at B. The difference is CD which (because the two figures are similar) is $(mv)\ (vt)/r$.

It was Newton who launched the great idea that some sort of force is required to make any such change in momentum; quantitatively, the force is the change in momentum in time t, which Figure 1 shows to be $(mv)(vt)/r$, divided by t. So the tension in the string must equal mv^2/r. What similar force keeps our Moon going round the Earth? Newton himself first wondered if it might be the familiar force of gravity mg, which pulls down towards the centre of the Earth any released object with acceleration g (amounting to a velocity increase of 9·8 m/s in every second).

Actually, the equation $mg = mv^2/r$ would give the Moon a velocity $v = \sqrt{(rg)}$, where $r = 380{,}000$ km is its distance from the centre of the Earth. However, it would then complete its orbit (a distance $2\pi r$) in a time $2\pi\sqrt{(r/g)}$—about half a day! So Newton had to think again . . . . He saw that the right answer (28 days) might be found if gravity fell off in strength with distance from the Earth's centre. This is where Galileo's four moons of Jupiter came in so useful. The time they took to go round was much bigger for those at greater distances r from Jupiter's centre, and it varied not as $\sqrt{r}$ but as $r\sqrt{r}$. This suggested that g may fall off in proportion to $1/r^2$ from its value at the surface, both for Jupiter and for the Earth.

It must have been exciting when Newton realized that on this assumption $2\pi\sqrt{(r/g)}$ for the Moon takes the value 28 days. Subsequent checks on Newton's 'universal inverse-square law of gravitation' have shown that it explains very accurately indeed all the motions within the solar system (including all the departures of orbits from exact circular shape).

What is involved, then, in getting craft into space? Above all, you must get the craft up to a speed $\sqrt{(rg)}$, which is 8 km/s at altitudes of a few hundreds of kilometres (at higher altitudes craft can be orbited at lower velocities, but the extra work required to lift them up there *more* than cancels the reduction in work due to this). Only at speeds as fast as 8 km/s will the craft 'stay up' at such altitudes; that is, go round and round in an orbit keeping clear of the Earth's atmosphere, which, at these speeds, would burn it up due to air friction.

Thus the craft must be lifted through the atmosphere, and is only then given a horizontal velocity of 8 km/s. Formidable task! A gun would be useless, since it imparts velocity while the craft is still in the atmosphere. So does an aircraft engine, which swallows vast quantities of oxygen from the air to 'burn' fuel in (that is, chemically combine oxygen and fuel), and so energize a jet which by throwing backward momentum away constantly imparts forward momentum to the craft.

The only practical solution is the rocket, which carries its own oxygen on board (usually in liquefied form) for chemical combination with fuel in a combustion chamber. Such a rocket can continue to provide acceleration *after* it has climbed out of the atmosphere, when indeed the main speeding-up must occur. Practical limitations on the chemical energy of fuels, and on the ratio (about 15 at best) of initial rocket mass to the mass remaining when all fuel is burnt, prevent the full 8 km/s velocity being reached with a single rocket. At least two are needed, with the bigger one dropping off when all its fuel is burnt and a second one ignited only then.

The spacecraft is stabilized and controlled by swivelling the rocket motors to provide a variable direction of thrust to correct automatically toppling of the craft sensed by gyroscopes within it, or to respond to radio-control signals from the ground. Multi-stage rockets can now be guided into circular orbits round the equator at a specially interesting altitude, about 30,000 km, for which the time taken for one revolution is the same (24 hours) as the spinning Earth takes. Communications satellites are particularly good in such an orbit, because they remain always over the same point on the Earth's surface, so that transmitting and receiving aerials on the ground can be permanently focused on them!

Problems

Readers who have not yet reached the age of 20 on 1 October 1968 are invited to submit solutions to 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.

1. Let G be the centroid of the acute-angled triangle ABC of circumradius R. Show that

$$AG^2+BG^2+CG^2 \geqslant 2R^2.$$

2. Let $f(x)$ be a polynomial of degree n with real coefficients and such that $f(x) \geqslant 0$ for all real x. Show that

$$f(x)+f'(x)+f''(x)+\ldots+f^{(n)}(x) \geqslant 0$$

for all real x.

3. Let $z_1,\ldots, z_n$ be complex numbers in a sector of the complex plane with the origin as its vertex and of angle θ, where $0 \leqslant \theta < \pi$. Establish the inequality

$$\left|\sum_{k=1}^{n} z_k\right| \geqslant \cos\tfrac{1}{2}\theta \sum_{k=1}^{n} |z_k|.$$

4. Show that the set of all positive integers can be partitioned into two subsets neither of which contains an infinite arithmetic progression.

5. A problem on Bayes' theorem is given at the end of the article on pp. 9–13. Find the respective probabilities that the red ball has come from one of the boxes A, B, or C.

6. The probability of a student's passing an examination is p, where $0 < p < 1$. It is assumed that his efforts in consecutive examinations are independent. Find, to 2 significant figures, the minimum value of p such that the probability of a student's passing in not more than 10 attempts is at least $\frac{1}{2}$.

Letters to the Editor

Dear Editor,

While I was visiting Stanford this summer, I saw two mathematical education films which I found very interesting. Both films seemed to me most suitable as educational aids for schools in Britain.

The first one, aimed at younger pupils (say second to fourth forms), presented a mathematical problem within a fanciful story called *Mr. Simplex Saves the Aspidistra*. The problem is of the chessboard parity type, and the viewer is taken through a natural process of problem solving. There follows an exploration of related problems and generalisations, and some questions are left unsolved for the viewer to do. Some remarks directed mainly to the teacher explain why such problems, although they have no immediate application, should be included in the scope of mathematics.

The second film, *What is Area?*, requires some knowledge of limits and is therefore suitable only for sixth forms, although maybe other pupils would gain something from it too. Area is defined for an arbitrary shape by considering the limits of the areas of inscribed and circumscribed polygons. The language used is that of set theory and analysis. Without being too formal, some of the concepts from Measure Theory are introduced. I feel that this film would give sixth formers a glimpse into the sort of mathematics they will study at university. The next film in the series is on integration.

I should like to know whether films of this sort have been made in Britain. If not, why not? Also, could these American films be made available here and distributed to schools, or be hired out by the L.E.A.'s? At least the Colleges of Education ought to be able to get hold of them.

Can you give me any further information, please.

Yours sincerely,

Richard Morton

(University of Manchester).

* * *

Dear Editor,

I should like to suggest that *Mathematical Spectrum* contains an Exchange and Mart for ideas and problems, theorems and conjectures sent in by amateur research mathematicians, including sixth formers and undergraduates.

The aim would be not so much to publish complete proofs of theorems but rather to help the flow of ideas by (i) linking those working on the same problem, (ii) giving references to standard results, (iii) indicating open and closed questions and associated conjectures (amateur and professional), (iv) outlining problems which have proved a fruitful source of class work, (v) giving results (not necessarily with details of proof) of recent amateur research, e.g., results obtained by exhaustive case by case examination.

Yours sincerely,
A. K. Austin
(University of Sheffield).

Mathematical Spectrum would be glad to publish any suitable ideas, problems, theorems and conjectures in the Letters to the Editor section.—*Editor*.

Notes on Contributors

Frank Benson is a Reader in Operational Research at the University of Southampton. He has taught at the University of Pennsylvania, Case Western Reserve University, and the University of Birmingham.

J. Gani is Professor of Statistics and Head of the Department of Probability and Statistics in the University of Sheffield. He has taught in universities in this country as well as in Australia and the U.S.A.

Jack Howlett is Director of the Atlas Computer Laboratory, an organisation under the Science Research Council, which provides computing services to research workers in all British universities.

G. T. Kneebone is Reader in the Foundations of Mathematics at Bedford College, University of London. He is author or joint author of several books on mathematical logic, geometry, and the theory of sets. He has, for some years, acted as external examiner in mathematics for the University of East Africa.

M. J. Lighthill FRS, is a Royal Society Research Professor and is at present at Imperial College, University of London. He is soon to take up the Lucasian Professorship of Mathematics in the University of Cambridge. Prior to his present position, he was for some years the Director of the Royal Aircraft Establishment at Farnborough.

L. Mirsky is a Reader in the Department of Pure Mathematics in the University of Sheffield. His interests include the theory of numbers, combinatorial analysis, and matrix algebra.

George Pólya is Professor Emeritus of Mathematics at Stanford University, California. He is a leading mathematical analyst and has written over 200 papers in research journals, as well as a number of books on the art of mathematical discovery.

R. J. Webster is a Lecturer in Pure Mathematics in the University of Sheffield, and has also taught in Summer Schools organised by the University of Toronto. His principal research interest is the study of convexity.

Contents

	1	Editorial
R. J. WEBSTER	3	An extremal problem in elementary geometry
F. BENSON	6	Where shall we build it?
J. GANI	9	Whodunit? Or the Reverend Mr Bayes FRS helps to decide
G. T. KNEEBONE	13	Mathematics and the physical world
	16	Professor Polya on intuition and learning
J. HOWLETT	16	Computers and their uses
L. MIRSKY	21	From rule of thumb to abstract structure
M. J. LIGHTHILL	24	Space
	29	Problems
	30	Letters to the Editor
	31	Notes on Contributors

PRICES (*postage included*)

	Individual annual subscription	Bulk annual subscriptions (5 or more copies)	Single issues
U.K. only	6/-	5/- each	3/6 each
Australia	$A1.10	$A0.90	$A0.60
U.S.A.	$1.20	$1.00	$0.70

Enquiries about rates and subscriptions should be directed to:

Oxford University Press, Education Department,
Walton Street, Oxford,

or to Oxford University Press Branches Overseas.