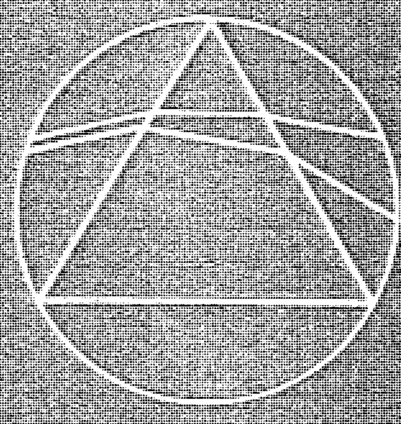
Mathematical Spectrum



Volume 4 1971/72 *Number 2*

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.

This number 2 completes Volume 4 of *Mathematical Spectrum* (1971/72); the first issue in the Volume was published in the Autumn of 1971.

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

Problems in the Use of Educational Tests

PAUL H. JACKSON

University College of Wales, Aberystwyth

The primary school teacher gives a spelling test. The Local Education Authority[1] conducts (or decides not to conduct) an 11-plus examination. The O- and A-level industry has an annual turnover of hundreds of thousands of pounds. University staffs devote hours of time, which could be more creatively spent, to the careful setting, scrutinising and marking of degree examinations. What is all this educational testing intended to achieve? And what are the difficulties in the way of achieving it?

1. Testing as measurement

The case of the primary school teacher is perhaps the most straightforward. Earlier in the week she has selected twenty words which, in her judgement, an educated child ought to be able to spell, and given them to her class to learn. The object of the test is to determine, for each child, the extent to which the learning task has been completed. Since all twenty words are judged to be a necessary part of a child's vocabulary, it is immaterial whether they are equally difficult to spell—a score of fifteen correct indicates that five words remain to be learnt. Also at the time of the test the child either does or does not know how to spell each of the twenty words. It seems, then, that the spelling test is a measuring instrument of unambiguous intent and perfect accuracy.

But is it? What do we say about the child who comes home in tears because she accidentally wrote the letter 'b' back to front, resulting in her 'table' being read as 'tadle' and marked wrong? Is she merely taking life too seriously? Or is she right in feeling that a chance occurrence, unrelated to the main purpose of the test, has resulted in her achievement being underrated (a serious injustice at her age)?

The fact that "error" in educational measurements results not only in inefficiency, but also in injustice to individuals, explains why the debate on the merits or demerits of various educational tests is conducted in a glare of publicity, while thousands of equally error-prone measurements in other contexts attract no public attention.

Of course, "error" may increase and not reduce one's score. Many adults, even highly educated ones, have a residual list of words which they frequently misspell, although from time to time they hit on the correct spelling. In these circumstances it would seem proper to say that they do not know how to spell the word but sometimes, by chance, appear to do so.

2. "Error" due to sampling

If we move from a weekly test on a specific list of words to an end-of-the-year spelling test, the problem of "error" looms larger. The object of the test can no longer sensibly be defined in terms of the particular list of words presented to the examinees: we are now concerned with 'spelling ability' in some more general sense. One formulation might be as follows:

During the year the class has been given 500 different words to learn. The percentage of these which a child can spell correctly will be called its 'spelling ability'.

If the end-of-year test contains 25 words, multiplying the number correct by 4 will give a measure of the pupil's apparent spelling ability. But clearly we now have an extra source of "error". A pupil's score will depend on *which* list of 25 words, chosen from the 500, is read out for him to spell. To be sure, good spellers will get most words from any list correct, and poor spellers will get most of them wrong. But it is not at all unlikely that a moderate speller might get 10 correct from one list of 25, and 15 correct from another, giving him apparent 'abilities' of 40% and 60% respectively. The discrepancy between these can only be ascribed to chance "error" resulting from the necessity of selecting a sample of words for the test.

How serious are the effects of this new source of "error" likely to be? We can get some idea by supposing that the 25 words of the test are a random sample from the 500 words learnt, chosen by a chance mechanism, such as writing each word on a slip of paper and drawing words lottery-style, which ensures that any collection of 25 is just as likely to be chosen as any other collection. This would be one way, not necessarily the best, of making the test representative of the year's work.

If we construct the test in this way, the number of words in it which a pupil knows how to spell correctly will be related to his 'spelling ability' by the binomial probability mechanism discussed by J. Kiefer in an earlier issue of this magazine (Volume 3, pp. 1–11). (Strictly speaking this is true only if we replace the slips of our word-lottery after each drawing and run the risk of including the same word twice. In the present context the difference between this and the more natural procedure of sampling without replacement can safely be ignored.) In the language of Kiefer's article, the purpose of the test might be to obtain a point estimate of the pupil's spelling ability.

Our immediate concern here, however, is with the extent to which "error" resulting from the choice of a particular set of 25 words may make the result of the test misleading. Figure 1 shows some aspects of the relation between a pupil's apparent ability X (four times his number correct) and his true ability T (Greek capital 'tau'). The bold line gives the expectation,

$$E(X|\mathrm{T}) = \sum Xp(X|\mathrm{T}),$$

of the apparent ability of a pupil whose true ability is T; here $p(X|\mathrm{T})$ denotes the binomial probability that a pupil with true ability T has apparent ability X. The dotted lines give, approximately, the boundaries within which there is a 50 : 50 chance that the apparent ability of a pupil with true ability T will fall. The continuous outer lines give similar boundaries with a 10 : 1 chance of enclosing his apparent ability.

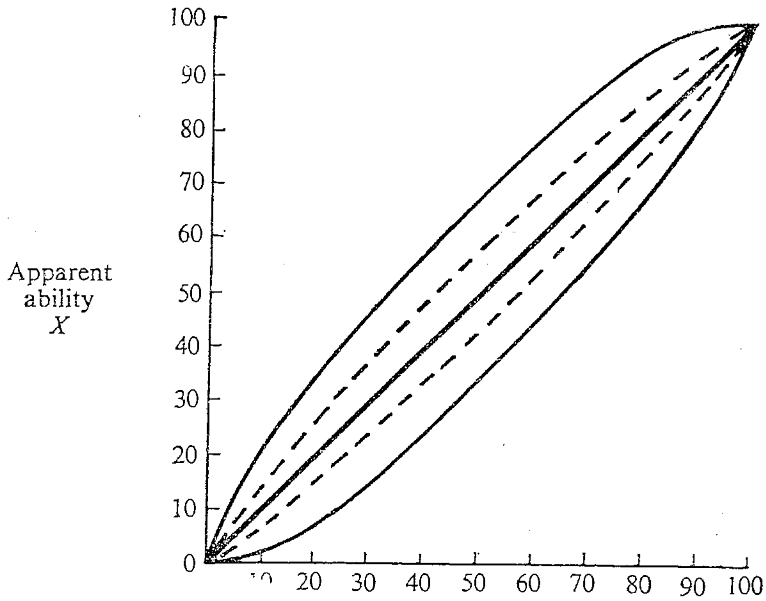


Figure 1. Relation between apparent and true ability.

Two features of the figure are worth noting. Firstly, the expectation of apparent ability X is equal to the true ability T. And secondly, over most of the ability range the effect of "error", as measured by the distance of the boundaries from the line giving the expectation, is similar whatever the true ability. Of course, at the ends of the range there is no error at all—a pupil who knows all 500 words will score 25 out of 25 on any test, and a pupil who knows none will score zero. This is sometimes referred to as the 'ceiling' and 'floor' effect.

Note also that the "error" due to sampling is quite likely to produce "injustice". A glance at the figure shows that it is not at all improbable that a pupil whose true ability is 45% may have a higher apparent ability than one whose true ability is 55%.

3. Other sources of "error"

The "error" due to sampling discussed in the previous section is, of course, additional to the accidental or irrelevant "error" mentioned previously. Even so, we have relied on two rather special features of spelling tests to reduce sources of "error". One of these is that spellings are unambiguously right or wrong (provided the test is firmly located on one side or other of the Atlantic!). Another is that it may be fair to take the percentage of the year's words known by the pupil *on the day of the test* as his true ability score.

But what about a test in English composition? Different teachers may mark the same essay differently. And the quality of a pupil's writing on Friday may well be different from the quality he would have produced in an essay on the same topic the following Monday. Which marking, and which test day, yield the pupil's true score? Indeed, what does 'true score' mean in this context?

4. A mathematical model for test scores

The way ahead involves a firm grasping of the nettle. For a variety of reasons beyond our control, a pupil's observed score X is not a fixed number characterising his ability, but varies from test to test, marker to marker, occasion to occasion. Very well then, let us study the probability distribution of X. (Of course, variation due to causes which are within our knowledge or control, such as one teacher being a known stingy marker, should be adjusted for.)

In the example of sampling error we saw that the observed score distribution was, very nearly, binomial. For more general sources of error we suppose any particular pupil's observed score X to have expectation T and variance σ_E^2, and call T the pupil's true score. Note that whereas in the spelling test we could frame a sensible definition of true ability without relation to observed score, and then *prove* as a consequence of the random sampling procedure that true score was equal to expected observed score, in the more general case we *define* true score in this way. Of course, the definition is chosen to be intuitively reasonable. For example, we can think of T as being the average score a student would obtain over many repetitions of the test (or an equivalent test).

The quantity

$$E = X - \mathrm{T},$$

the difference between observed score and true score, is called the error score. Since for any particular pupil T is a fixed number, the variance of X is the same as the variance of E, which explains the use of the symbol σ_E^2 for the variance of the distribution of X. This variance is, in fact, simply the expectation of the square of the error score and is usually called the error score variance or error variance.

We commonly assume that the error variance is the same for all pupils: in particular we assume it does not depend on T. In discussing sampling error we saw that this was reasonable provided we were not getting 'ceiling' or 'floor' effects. By arranging the difficulty level of a test so that hardly any examinees get all or none of the questions right, this requirement can often be met.

How far has this mathematical modelling advanced us? If we have only one observed score for each pupil, not one inch! But suppose we could somehow get two observed scores X_1 and X_2 for each pupil. Then the difference $X_1 - X_2$ would give us some idea of the size of the error scores involved. Assuming a common error variance for all pupils, we should have a collection of 'difference scores', $X_1 - X_2$, one for each pupil, from which we could make a fairly good estimate of σ_E^2. This in turn would allow us to answer questions such as: if pupil A has averaged 45% and pupil B 50% on the two tests, how likely is it that A is in fact the better pupil (has the higher true score)?

Of course, this procedure will make sense only if the two tests are measuring the same thing in the same way and on the same scale. They should be equally difficult and neither inherently more subject to error than the other. More formally, we call tests parallel or equivalent if each pupil has the same true score and error variance on each of the tests. Parallel tests can be constructed by careful matching of items, or sometimes in the case of tests consisting of large numbers of items, by treating the odd- and even-numbered items as two separate tests.

5. The reliability of tests

Once we have an estimate of the error variance, we can begin to discuss questions about the accuracy or reliability of tests. For example, do multiple-choice papers measure true score more or less accurately than essay-type papers? (A separate, and important, question is whether they measure the same thing, that is, whether true score on a multiple-choice paper reflects the same educational characteristic of a pupil as true score on an essay paper.) But now we run into another snag. The range of marks for an educational test is arbitrary. We can mark from 0–10, 0–100, 20–80, or any other range we choose, and the value of σ_E^2 will change accordingly. Nor is the nominal range of marks of much use. A mathematics paper and an English essay may both be scored as percentages. But while it would not be extraordinary to find scores around zero and one hundred on the mathematics paper, it would be most unusual to find such extreme scores for the essay.

Another grasping of the nettle is called for. A characteristic of the group of pupils we are testing is that ability varies from pupil to pupil; formally, pupils have different true scores. Let us denote by σ_{T}^2 the variance of pupils' true scores. A decision to change the range in which scores are reported is equivalent to making a linear transformation, $Y = aX+b$, of the observed scores. The effect of this is to multiply both σ_E^2 and σ_{T}^2 by a^2. Hence any measure of accuracy or reliability which is essentially a *ratio* of these two quantities will be independent of the range chosen. Another way of putting this is that in order to be able to compare error score variances for different tests we must first ensure (by changing the range) that they have the same true score variance.

The particular ratio commonly used in educational measurement is the reliability,

$$\rho = \frac{\sigma_{\mathrm{T}}^2}{\sigma_{\mathrm{T}}^2 + \sigma_E^2},$$

which takes the value 1 when the measurement is error-free, and tends to zero when the error component of X is completely swamping any true score variation present. The reason for the choice of this particular ratio is that it can be interpreted as a correlation coefficient. In fact it is the correlation between the scores on two parallel tests, and so is a natural measure of the 'repeatability' of a score. (In other contexts the signal-to-noise ratio, σ_T^2/σ_E^2 serves essentially the same purpose.)

Of course, since we can never observe T, we can never know σ_T^2 precisely. But we can estimate it. A consequence of our model is that the overall variance of observed scores, σ_X^2, can be expressed as the sum of two components

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2.$$

We can estimate σ_X^2 from the observed variation among the scores recorded for the test(s), and we have already seen how to estimate σ_E^2 using parallel tests. By subtraction we arrive at an estimate of σ_T^2. Plugging these values back in the formula for ρ we get an estimate of the reliability of the test.

It is worth repeating that the reliability ρ measures the accuracy or repeatability with which the observed score X indicates the value of the true score T. It tells us nothing whatever about the meaning, usefulness, or relevance to our problem of T itself.

6. Estimating an individual's true score

If the only information we have is the test score X of an individual candidate, we can do no better than estimate his true score by

$$\hat{T} = X.$$

If, however, we have a group of candidates from a particular institution, we can consider using the linear regression equation of T on X to make our estimates. If the reliability ρ and the mean true score μ_T for that institution were known, this would lead to the estimate

$$T^* = (1-\rho)\,\mu_T + \rho X.$$

In practice, ρ will have to be estimated by the method described previously, and μ_T by $\bar{X}$, the mean score of the candidates from the institution.

The regression equation gives 'better' estimates of true score than do the individual observed scores, in the following sense. If we use T* as our estimate for a candidate whose true score is actually T, the error in our estimate is (T^*-T). An indication of the accuracy or otherwise of our estimation procedure is given by $E(T^*-T)^2$, the expected squared error of estimation. It can be shown that this is less than $E(\hat{T}-T)^2$; indeed, it is a property of the regression line that $E(T^*-T)^2$ is less than the expected squared error of estimation of any other estimate which is a linear function of X. Thus from the tester's point of view, its use improves efficiency.

But what of the candidate's point of view? Since $0 \leqslant \rho \leqslant 1$, the estimate T* will be between the candidate's observed score X and the mean μ_T of the institution with which he is associated. If the reliability of the test is high, T* will not be very different from X. But when the reliability is low the estimate will move close to μ_T.

This 'going back' towards the group mean is the phenomenon which gave regression lines their name.

A candidate with an above-average test score will have it reduced by the regression estimate: is he entitled to feel aggrieved? More seriously still, if two candidates with the same test score X belong to different institutions having different mean scores, their estimated true scores will be different: is this fair?

The source, though not the resolution, of this conflict between efficiency and fairness is easily found. Prior to seeing the test paper, the only information available to the tester is the institution from which the candidate comes: he is therefore entitled to treat the candidate as randomly sampled from that institution, and he will increase the efficiency of his estimation procedure by doing so. The candidate, however, is unlikely to regard himself in this light, and if he regards himself as above average for his institution, he will resent having his test score regressed towards the institution mean.

7. Testing as classification

If we enquire what is the purpose of 11-plus examinations, the answer cannot be framed in the terms we have used so far. The real purpose is not to measure anything, but to classify pupils as 'grammar' or 'modern' ('grammar', 'technical' or 'modern' where the 1944 Act has been fully implemented). Naturally, if we base our classification on a measurement which has low reliability, we must expect a lot of misclassification. This is one reason for giving considerable weight to 'Intelligence Tests', which are typically highly reliable, more so than Arithmetic tests and much more so than English tests.

But reliability itself is no guide to what should be included in an 11-plus test. A pupil's height can be measured with nearly 100% reliability, but this does not mean one should allocate tall pupils to grammar school and short ones to modern school!

We are compelled to ask radical questions about the purpose of the classification. What does one go to grammar school for? To get some A-levels? To meet the boy/girl one eventually marries? To develop one's interest in art, music, drama, or sport? When I sit at my desk in the university, I naturally think of A-levels. But when I sit on a rock by the sea shore with the tide lapping in around me and the sun sinking slowly behind the horizon, I know with an inner certainty that one's 'scores' on the other two questions are much more important.

My excuse for intruding philosophical thoughts into a mathematical paper is that they emphasise the obvious, but often forgotten, fact that one cannot discuss what is a good statistical procedure without pinning down what it is desired to achieve; this in turn usually involves making explicit one's value judgements. If allocation to grammar school is to be based solely on potential for academic achievement (as measured by O- and A-level successes), then this must be justified either on the grounds that this is the *only* area of educational activity for which the grammar school provides more appropriate facilities than the modern school, or that this area is of overriding importance.

8. Measurement as prediction

If we accept that 11-plus classification is intended to assign to grammar schools those children who will perform best at O-level, or that university selection should lead to the acceptance of those students who will get the best degrees, we can view the statistical problem involved as one of prediction (e.g., of degree result from A-level result). Moreover, the prediction model has many uses besides selection. For example, if we are able to tell a student that at institution A he is likely to get a bare pass but at institution B he is likely to emerge with flying colours, this information may be of great help to him in choosing his course of action. Just *how* he uses it is for him to decide. If he has a strong preference for institution A, the crucial part of the information *for him* might be that he has a fair chance of getting a pass of some sort there.

Before we can consider prediction of one thing on the basis of another, we must agree how to measure the two 'things'. One system used to reduce A-level results to a single number is to allocate 5 points for a Grade A pass (the highest), 4 for Grade B, 3 for C, 2 for D, 1 for E (the lowest pass), and then total up the points. This innocent-looking 'coding' is anything but uncontroversial. Is an A-level result consisting of two grades C exactly as good as one consisting of three grades D? Better? Worse?

In a short article we must ignore problems like this and press on. Let X denote total A-level points. Similarly, for degree result let Y be the student's average mark on his final examination papers. We would like to be able to write a prediction equation

$$Y = f(X)$$

which would enable us to predict our criterion Y (degree result) when the value of the predictor X (A-level result) is known. Of course, no one expects the prediction to be exact. Not only are the variables X and Y both subject to errors of measurement in the sense discussed previously, but the relation between Y and X is clearly affected by how a student uses his time after he arrives at university. An indication of how far any particular student's result is likely to be from the value given by the prediction equation would therefore be useful.

The data we can use to decide on a prediction equation are the records of students who have already completed their degrees. Unless one is prepared to make the sweeping assumption that the same prediction equation will be appropriate for all universities, one will have to consider each university separately. Indeed, a further narrowing down to separate faculties of Arts, Science, etc. will usually be necessary. Add to this the fact that rapid changes in secondary and higher education are making data obsolete within a few years, and our data-base is beginning to look a bit scanty.

If the data available were extensive, we might proceed as follows: take all the students in the relevant records with a particular A-level score X and find their average degree score, say $\bar{Y}_X$; plot a graph of $\bar{Y}_X$ against X; with any luck at all the plotted points will lie on a smooth curve whose equation we can take to be our prediction equation.

Unfortunately, the data available are unlikely to be anywhere near adequate to making this procedure workable: the result of using it for a set of 236 science students is shown in Figure 2. Hence, in the absence of any strong evidence to the contrary, we suppose that the simplest plausible equation, the straight line

$$Y = a + bX,$$

will serve adequately for prediction, and seek the "best" straight line through the available data points (X, Y).

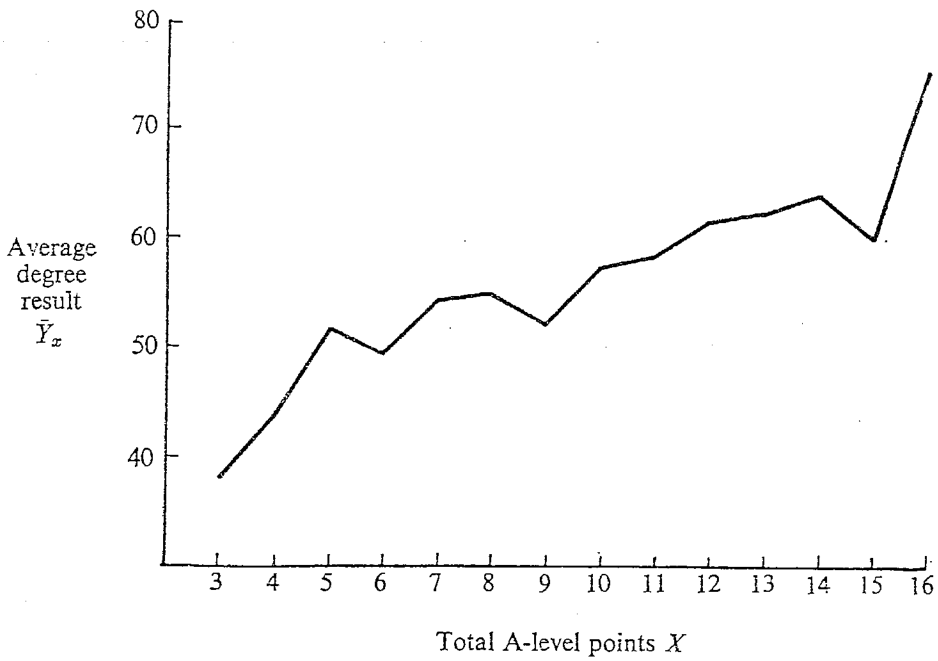


Figure 2. Average degree result of students according to total A-level points.

Suppose there are N such points, and the means of the predictor and criterion scores are $\bar{X}$ and $\bar{Y}$ respectively.

Let

$$S_{XX} = \Sigma(X-\bar{X})^2, \quad S_{YY} = \Sigma(Y-\bar{Y})^2, \quad S_{XY} = \Sigma(X-\bar{X})(Y-\bar{Y}),$$

all sums being taken over the complete set of data points. The error of prediction made when using the line for a student with predictor score X and criterion score Y is $Y-(a+bX)$, which may be positive or negative. The sum of squares of prediction errors provides a measure of overall inaccuracy of prediction, and the usual "best" line, the regression line of Y on X, is chosen to minimise this sum,

$$\begin{aligned}\Sigma(Y-a-bX)^2 &= \Sigma[\{(Y-\bar{Y})-b(X-\bar{X})\}+(\bar{Y}-a-b\bar{X})]^2 \\ &= \Sigma\{(Y-\bar{Y})-b(X-\bar{X})\}^2+N(\bar{Y}-a-b\bar{X})^2,\end{aligned}$$

the sum of cross-products being zero, so that

$$\sum(Y-a-bX)^2 = b^2 S_{XX} - 2bS_{XY} + S_{YY} + N(\bar{Y}-a-b\bar{X})^2$$

$$= S_{XX}\left(b-\frac{S_{XY}}{S_{XX}}\right)^2 + S_{YY} - \frac{S_{XY}^2}{S_{XX}} + N(\bar{Y}-a-b\bar{X})^2,$$

on completing the square.

Since the first and last terms cannot be negative, minimisation is achieved by setting

$$b = \frac{S_{XY}}{S_{XX}}, \quad a = \bar{Y}-b\bar{X},$$

making these two terms zero.

Using the prediction line obtained in this way, the sum of squared errors of prediction is just

$$S_{YY} - \frac{S_{XY}^2}{S_{XX}},$$

and on dividing this by N (some people prefer $N-2$) we obtain an estimate of the variance of the error involved in any prediction. We noted earlier that some such indication of the size of error involved would enhance the usefulness of the prediction equation.

How useful is that equation anyway? The estimated error variance is not itself an answer to this question since it depends upon the scale of measurement for Y. As in the case of reliability discussed earlier, we overcome this problem by finding a scale-free ratio.

If the information contained in the predictor scores X were not available, the best prediction we could make would be that any student's criterion score would be $\bar{Y}$, the mean of our sample data. The corresponding error sum of squares would be

$$\sum(Y-\bar{Y})^2 = S_{YY}.$$

By using our knowledge of X, we have reduced this sum by a fraction,

$$\frac{1}{S_{YY}}\left\{S_{YY} - \left(S_{YY} - \frac{S_{XY}^2}{S_{XX}}\right)\right\} = \frac{S_{XY}^2}{S_{XX}S_{YY}}.$$

The square root of this fraction,

$$r = \frac{S_{XY}}{\sqrt{(S_{XX}S_{YY})}},$$

is the correlation coefficient between X and Y. The above derivation shows that it is a scale-free measure of the usefulness of the prediction equation. In educational testing it is spoken of as the *validity* of the test being used for prediction. Evidently a test has many validities, one for each predictive purpose to which it might be put. The calculation of the validities of several tests for one particular purpose enables us to decide which will be most useful for that purpose.

9. Conclusion

This article has examined two fundamental ideas in the theory of educational tests: the reliability of a test is concerned with the question "Does this test measure *something* in a dependable manner?" and is of interest when testing is viewed as measurement; the validity of a test is an answer to the question "Does this test measure what I want to know?" and is of interest when tests are used for prediction. The following list of additional questions we might wish to answer indicates some of the lines along which the theory of testing is being developed.

Many tests clearly measure a combination of abilities, skills and achievements: is it possible to separate out the respective contributions to the total score?

Are there contexts in which one can increase efficiency by noting, say, to what institution a candidate belongs, without at the same time appearing to be unfair to some candidates?

Are there better ways of using A-level results for prediction purposes than simply combining them into a single 'A-level score'?

If a series of tests is being conducted for a specific predictive purpose, what is the optimal allocation of the available testing time between tests of different types (or between different examination subjects)?

Suggestions for further reading

F. M. Lord and M. R. Novick, *Statistical Theories of Mental Test Scores* (Addison-Wesley, New York, 1968).

[1] For the sake of concreteness, this article is set in the context of the British school system. This is administered by Local Education Authorities (LEAs), corresponding to the main administrative divisions of the country—counties, cities, etc. The 1944 Education Act, which brought the LEAs into being, provided for three types of secondary school: Grammar Schools for the academically inclined; Technical Schools for those with marked practical abilities; and Modern Schools for the majority. (Relatively few Technical Schools have, in fact, been established.) Allocation to the different types of school is/was made on the basis of the 11-plus examination, named from the typical age of pupils taking it. The 11-plus examination includes, at the discretion of the LEA, an intelligence test, tests of English and Arithmetic, and a report from the head of the pupil's primary school. Since 1944, comprehensive (unsegregated) secondary schools have been established, initially as experiments, but more recently as the accepted policy of several LEAs. This development has been a focus of social and political controversy, in which the reliability of the 11-plus examination as an allocation instrument has been sharply questioned.

Grammar school pupils and their equivalents are examined at around age 16 for the General Certificate of Education, Ordinary Level—"the O-level". This serves both as a passport to the final two years of school for those who stay on after 16, and as a leaving certificate for those who do not. The former take, at around age 18, the Advanced Level examination of the General Certificate of Education—"the A-level". Selection for university entrance is made primarily on the results of the A-level examination.

Profits without Tears

D. W. SHARPE

University of Sheffield

Your Auntie Ethel has just left you a considerable sum of money, and you decide to use it to set yourself up in business, so you buy a small factory which makes just two products, baffin pins and gudgeon pins. (For the purposes of this article, the exact nature and the many uses of these types of pin are irrelevant.) But you have been studying economics at A level, and you are keen to make your fortune out of your baffin and gudgeon pin factory, so you are not content to manufacture these merely as the fancy takes you; you want to maximize your profits.

The process whereby these pins are made is a particularly simple one, and uses only two raw materials which, for the purposes of industrial secrecy, we shall label with the enigmatic symbols A and B. Table 1 shows the number of hundredweights of A and B required to manufacture one hundredweight of baffin respectively gudgeon pins; Table 2 shows the profit in £'s that you make on each hundredweight of each type of pin made; Table 3 shows the availability in hundredweights of each of the raw materials each day. (You observe that the process is rather wasteful in the use of raw materials, but that is a matter for

TABLE 1

	baffin	gudgeon
A	1	2
B	3	1

TABLE 2

baffin	gudgeon
40	30

TABLE 3

A	B
2	3

your research laboratory.) Your problem is to determine how many hundredweights of each type of pin should be made each day to maximize the profits. (You already know that the demand is such that you will have no difficulty in marketing all the pins that you manufacture.)

Suppose you make x cwt. of baffin pins and y cwt. of gudgeon pins each day. Then your daily profit in £'s will be

$$40x+30y. \tag{1}$$

You are, however, limited by the availability of your raw materials. From Table 1, you will need $(x+2y)$ cwt. of A and $(3x+y)$ cwt. of B, so that, from Table 3,

$$\left.\begin{aligned} x+2y &\leqslant 2, \\ 3x+y &\leqslant 3. \end{aligned}\right\} \tag{2}$$

Thus you must maximize (1) subject to the restrictions (2). This is easily solved graphically. The shaded region S in Figure 1 shows all points (x, y) which satisfy the inequalities (2). (You insist that $x, y \geqslant 0$, as it is rather difficult to make a negative quantity of either variety of pin.) You must determine which point or points (x, y) of S will maximize the expression (1).

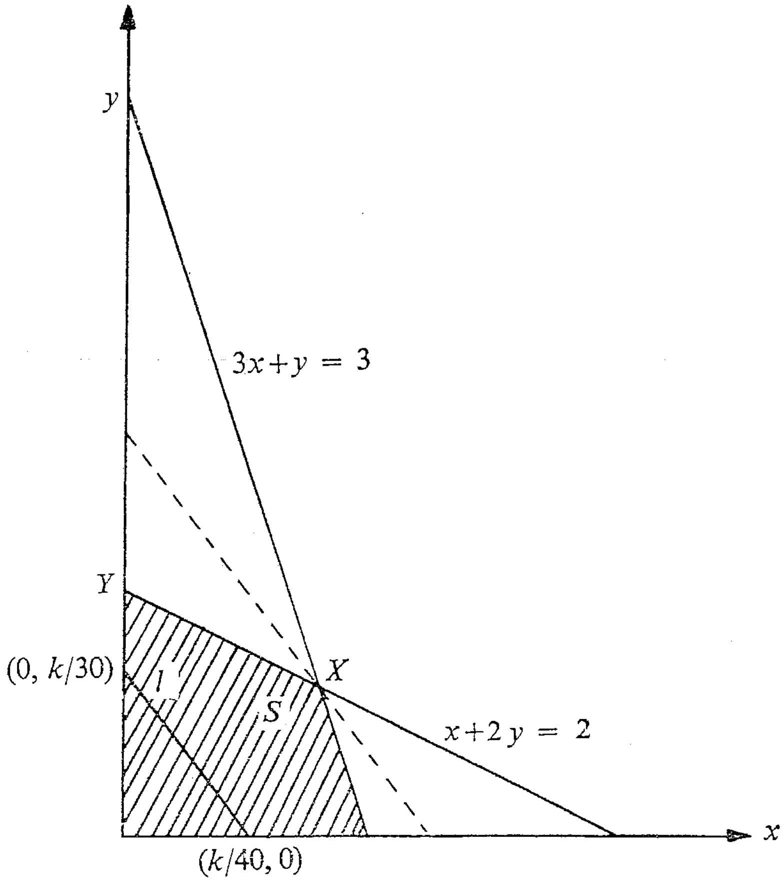


Figure 1

A little simple mathematics is called for at this point. The straight line l with equation

$$40x+30y = k$$

meets the x-axis in the point $(k/40, 0)$ and the y-axis in the point $(0, k/30)$. It should be clear from this that, the further is the line l from the origin, the larger is the value of k. Now the line is furthest from the origin when it passes through the point X on the diagram. (Of course, the whole point of the exercise is lost if l is allowed to move out of the region S altogether.) Thus the expression (1) is maximized at the single point $X = (\frac{4}{5}, \frac{3}{5})$. This means that you should manufacture $\frac{4}{5}$ cwt. of baffin pins and $\frac{3}{5}$ cwt. of gudgeon pins each day, and your maximum daily profit will be

$$£[40(\tfrac{4}{5})+30(\tfrac{3}{5})] = £48.$$

The mathematics is now over.

Suppose now that, because of increased imports from Hong Kong, it is necessary to reduce the price of your baffin pins in order to remain competitive, so much so that your profit on one hundredweight of baffin pins is reduced from

£40 to £10. Your daily profit function (1) is now

$$x+3y$$

and this is maximized at the point $Y=(0,1)$. In other words, you would do best to cease production of baffin pins altogether, and concentrate on gudgeon pins.

It is clear that this presents a somewhat oversimplified account of the manufacture of baffin and gudgeon pins. For example, the BGPWU (the Baffin and Gudgeon Pin Workers' Union) may be a little aggrieved to discover that you seem to have made no provision to pay your workers any wages. You will also have to take into account the depreciation of your machinery, distribution costs, etc. It does, however, illustrate a type of problem which is of considerable importance in economics. This is a *linear programming* problem. In general, we must find $x_1, x_2, \ldots, x_n \geqslant 0$ which maximize the linear expression

$$c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

subject to the linear inequalities

$$\begin{aligned}
a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n &\leqslant b_1 \\
a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n &\leqslant b_2 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots& \\
a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n &\leqslant b_m.
\end{aligned}$$

Here, the a_{ij}, b_i, c_j are real numbers. To solve this graphically requires n-dimensional graph paper, otherwise known as a computer. The method normally used is known as the *simplex method*, which is fortunately too involved to be described here.

Linear programming is a very recent branch of mathematics, and was first put on a secure foundation about 25 years ago by several mathematicians in America, notably J. von Neumann and G. B. Dantzig. It was originally devised to plan the movement of convoys across the Atlantic in the 1939/45 war. One of its early uses was in the planning of the Berlin air lift. The theory of linear programming employs various branches of abstract mathematics, in particular the theory of vector spaces and the theory of convex sets.

The moral of this seems to be that, before you can operate your baffin and gudgeon pin factory successfully, you need to know all about the latest branches of abstract mathematics; or else employ a mathematician.

Further reading

Mathematics in Management by Albert Battersby (Pelican, 1966) is an interesting popular book which describes many applications of mathematics in economics.

The Language of Geometry[1]

R. L. E. SCHWARZENBERGER

University of Warwick

Professional mathematicians have many differing answers to the question 'What is mathematics?' Some say the application of numerical calculations or logical reasoning; some the study of axiomatic systems; others are reduced to statements like 'Mathematics is what mathematicians do' or 'Mathematics is what appears in mathematics syllabuses'. My own attitude, which I share with many of my colleagues, is simply that mathematics is a language. Like English, or Latin, or Chinese, there are certain concepts for which mathematics is particularly well suited: it would be as foolish to attempt to write a love poem in the language of mathematics as to prove the Fundamental Theorem of Algebra using the English language. I do not make any attempt to defend this attitude in this article, but assume that mathematics is a language in which anyone, whether expert or non-expert, may attempt to communicate.

Now, just as in English we have a choice of whether to communicate in poetry or prose, whether to use arguments by deduction or by analogy, whether to use language which stimulates visual images or emotional responses; so in mathematics we have a choice of methods. A very common choice is between an axiomatic or algebraic method (involving calculation, formulae, and formal deductive logic) and a geometric method (involving spatial intuition, diagrams, and less rigorous argument). We tend to assume that, because the rigour of the argument is the most important test of a piece of mathematics, algebraic methods are automatically superior to geometric methods. But, if our attitude is that mathematics is a language, then there is another test which must be applied: which method communicates the truth in question with the greatest force and conviction?

I believe that it is often the more geometric method which communicates a mathematical truth most convincingly. There are examples of this all over mathematics (and also, incidentally, in English where it is often the argument or poem which carries a visual image which has the greatest impact) but I have space only for three. Those which I have chosen concern Pythagoras' Theorem about right-angled triangles, the solution of differential equations, and the classification of crystal lattices; I hope these are sufficiently varied to encourage the reader to look for similar examples from his own experience.

Pythagoras' Theorem

You feel an urgent wish to communicate the truth of this theorem, which of the methods developed during the last 2,500 years will you choose? Here are three.

[1] Based on a lecture given in July 1971 at a Mathematics Conference for Sixth-Formers at Bath University of Technology.

(i) *Coordinate geometry.* We begin by *defining* the distance P_1P_2 between points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ in the (x, y)-plane by

$$P_1P_2^2 = (x_1-x_2)^2+(y_1-y_2)^2.$$

Equally, if O is the point $(0,0)$, we *define* P_1OP_2 to be a right angle if

$$x_1x_2+y_1y_2 = 0.$$

With these definitions it follows that, if P_1OP_2 is a right angle, then

$$P_1P_2^2 = (x_1-x_2)^2+(y_1-y_2)^2 = (x_1^2+y_1^2)+(x_2^2+y_2^2) = OP_1^2+OP_2^2.$$

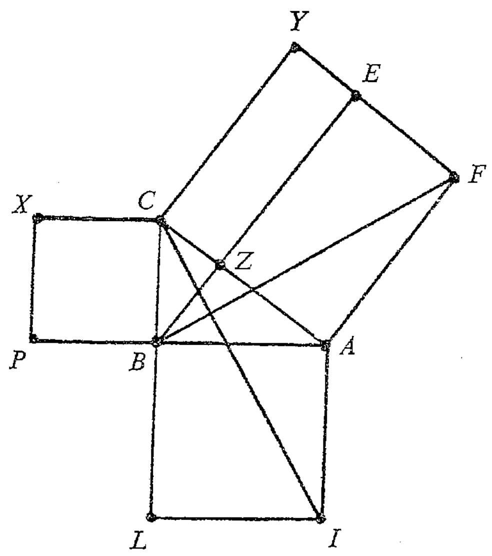
This argument is axiomatic, rigorous, and algebraic, but it is hard to claim that it carries conviction: one is left unsure whether Pythagoras' Theorem is a triviality or whether it is somewhere buried in the definitions. In fact the definitions of 'distance' and 'right angle' have been framed in such a way as to make the theorem nothing but a trivial algebraic identity.

(ii) *Euclid's proof.* This is the common name for the proof in which one considers a right-angled triangle ABC with squares constructed as in the first diagram. The result then follows from previously proved facts about areas of rectangles and triangles by a series of steps:

$$AB^2 = \text{Area}(ABLI) = 2\ \text{Area}(ACI) = 2\ \text{Area}(BAF) = \text{Area}(AFEZ),$$

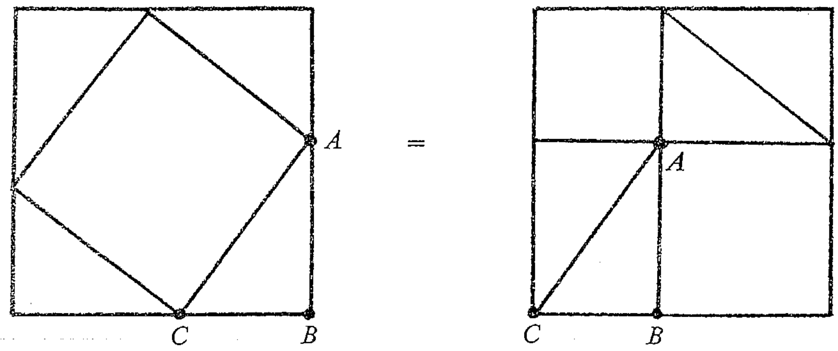
$$BC^2 = \text{Area}(BCXP) = 2\ \text{Area}(ACX) = 2\ \text{Area}(BYC) = \text{Area}(CZEY),$$

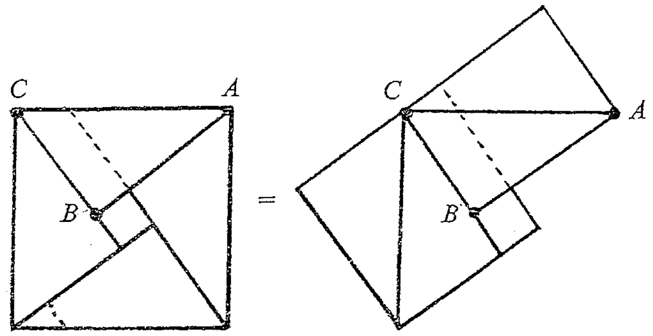
$$\therefore\quad AB^2+BC^2 = \text{Area}(AFEZ)+\text{Area}(CZEY) = \text{Area}(CAFY) = AC^2.$$



The proof remains rigorous, given Euclid's axioms, but it would be very difficult to follow without an accompanying diagram. For a person able to keep track of the diagram it carries conviction.

(iii) *Chinese proof.* By this I mean a proof which displays the formula by direct dissection of area. Such proofs can be found in Indian and Greek literature as well as Chinese. Two examples are illustrated: both convey the factual position with complete conviction but there are naturally slight doubts about the rigour of an argument which is so geometric that no words or symbols are needed.





Differential equations

You are asked to solve the equation

$$\frac{dy}{dx} = \frac{1}{x}$$

or perhaps to find the solution for which $y = 1$ when $x = 1$. Here are two methods of communicating the answer. One is tempted to add that the first method is perhaps more suitable for communication to an A-level examiner, the second for communication to a friend.

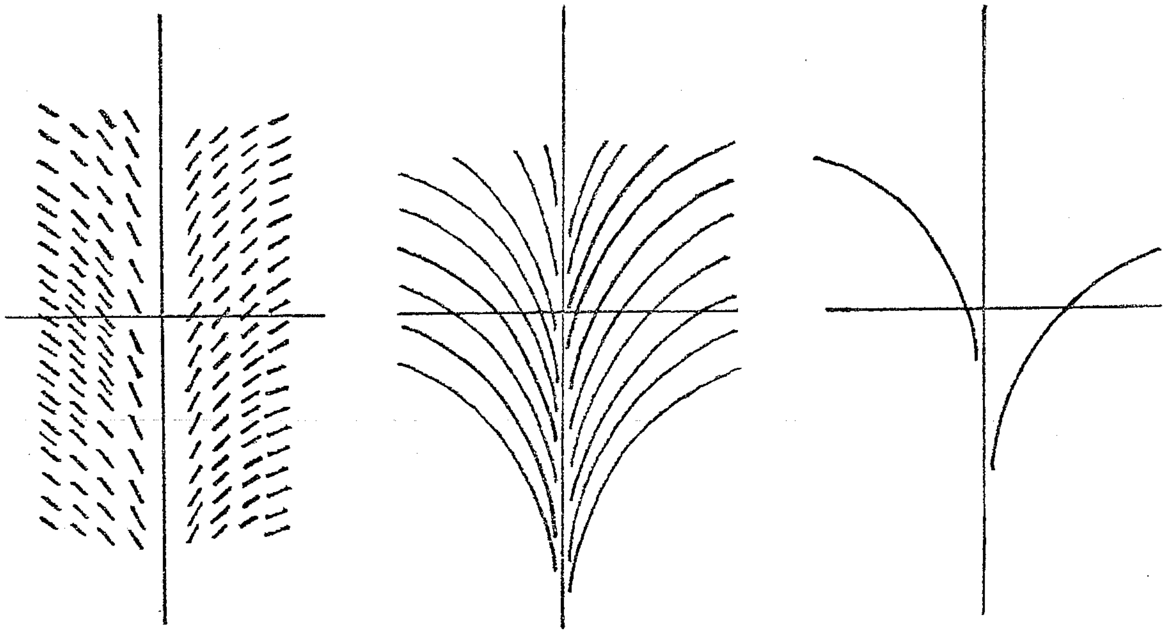
(i) *Calculation.* It is well known that the derivative of $\log|x|$ is $1/x$. Therefore every solution has the form

$$y = C + \log|x|$$

and the fact that $y = 1$ when $x = 1$ implies that $C = 1$. This is algebraic and straightforward, and it carries conviction provided you remember the theorem about arbitrary constants C correctly. It is however false because we have in fact not quoted it correctly!

(ii) *Plotting slopes.* Plot on a diagram of the (x, y)-plane what the equation means. At each point (x, y) with $x \neq 0$ the slope of a solution is $1/x$, at each point

$(0, y)$ the slope is infinite. Now draw curves to fit these slopes. We see that a typical solution consists of two curves:



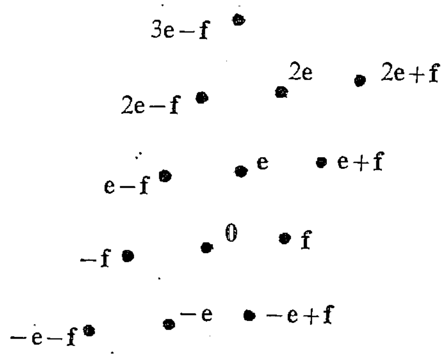
$$y = a + \log x \quad \text{when } x > 0,$$

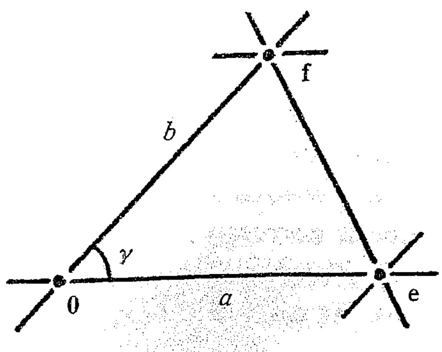
$$y = b + \log(-x) \quad \text{when } x < 0.$$

The fact that $y = 1$ when $x = 1$ implies that $a = 1$, but imposes no restriction whatsoever on the arbitrary constant b. In this case our use of a geometric diagram has the effect of preventing misquotations of the theorem about arbitrary constants, which should read: given a first order differential equation defined on a single interval of the real line, any two solutions differ by a constant.

Crystal lattices

So far the examples given have shown, first that a geometric argument can be more convincing than an algebraic argument, and second that a geometric method can assist in avoiding mistakes in an algebraic argument. The next example is a case in which a geometric method actually leads to a different result from the corresponding algebraic method. This is not to say that one method is wrong: but rather that the two methods deal with a problem in a slightly different manner.



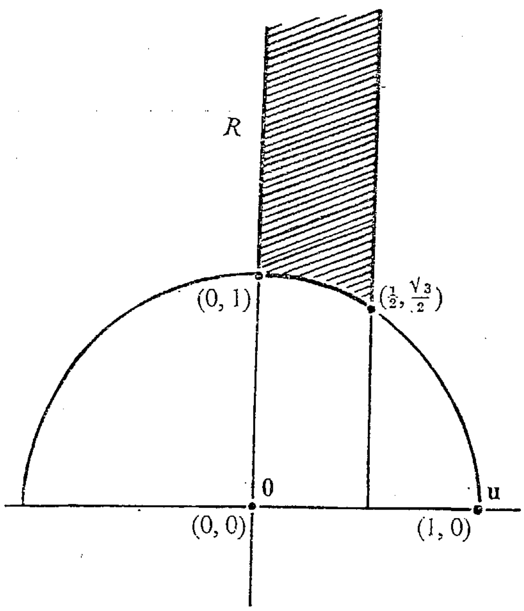


(i) *Algebraic treatment of plane lattices.* A plane lattice is an array in the plane of points of the form $m\mathbf{e}+n\mathbf{f}$ where $\mathbf{e}$ and $\mathbf{f}$ are two 'basis' vectors in the plane and m and n are integers. How many types of plane lattice exist? From the algebraist's, and crystallographer's, point of view the natural way to distinguish different types of lattice is to list them according to the symmetries of the lattice as follows.

Lattice	Restrictions	Symmetries of the lattice
Parallelogram	—, —	Rotation through 180°
Rectangle	—, $\gamma = 90°$	Perpendicular reflections in the lines of the basis vectors $\mathbf{e}$, $\mathbf{f}$, and rotation through 180°
Diamond	$a = b$, —	Perpendicular reflections in the lines of the bisectors $\mathbf{e}\pm\mathbf{f}$, and rotation through 180°
Square	$a = b$, $\gamma = 90°$	Reflections in lines of $\mathbf{e}$, $\mathbf{f}$, $\mathbf{e}\pm\mathbf{f}$, and rotations through multiples of 90°
Hexagon[1]	$a = b$, $\gamma = 60°$	Reflections in lines of $\mathbf{e}$, $\mathbf{f}$, $\mathbf{e}\pm\mathbf{f}$, and rotations through multiples of 60°

[1] The six shortest non-zero vectors $\pm\mathbf{e}$, $\pm\mathbf{f}$, $\mathbf{e}-\mathbf{f}$, $\mathbf{f}-\mathbf{e}$ form a hexagon with centre $\mathbf{0}$. This lattice could equally well be called 'equilateral triangle'.

These 5 types are often referred to as Bravais Lattices in honour of the French crystallographer Bravais who listed the corresponding lattices in three-dimensional space: in that case there are 14 types of lattice all of which occur in nature as the regular arrays of molecules in crystals.



(ii) *Geometric treatment of plane lattices.* Instead of listing symmetries, think about the possible positions of the basis vectors $\mathbf{e}$ and $\mathbf{f}$. It is natural to regard two lattices as equivalent if they differ only by a combination of rotations, reflections or changes of scale: thus all square lattices are equivalent regardless of the size of the squares or the angle at which they are placed in the plane. This means that we can choose basis vectors **u** and **v** (not necessarily the same as **e** and **f** above) in such a way that **u** is the shortest vector in the lattice. We may assume that **u** is the point $(1, 0)$ and that **v** is the next shortest vector which is not a multiple of **u**. What are the possible positions for **v**? Because **u** is the shortest vector, **v** lies outside the circle of radius 1 and centre the origin. Because **v** is the next shortest, **v** lies in the region of points (x, y) with $-\frac{1}{2} \leqslant x \leqslant \frac{1}{2}$. Reflections in the lines $x = 0$ and $y = 0$ can, if necessary, ensure that **v** lies in the region $x \geqslant 0$, $y \geqslant 0$. At this stage there is a region R of possible positions for **v** and there are essentially six different types of position; for the 'thin' diamond $\mathbf{e} = \mathbf{v} - \mathbf{u}$ and $\mathbf{f} = \mathbf{v}$, but in all other cases it is possible to take $\mathbf{e} = \mathbf{u}$ and $\mathbf{f} = \mathbf{v}$.

Position of **v** in region R	Type of lattice
v in interior	Parallelogram
v on edge $x = 0$	Rectangle ($\gamma = 90°$)
v on edge $x = \frac{1}{2}$	'Thin' diamond ($\gamma < 60°$)
v on curved edge	'Fat' diamond ($60° < \gamma < 90°$)
v at corner $(0, 1)$	Square ($\gamma = 90°$)
v at corner $(\frac{1}{2}, \frac{1}{2}\sqrt{3})$	Hexagon ($\gamma = 60°$)

There are now 6 distinct types of lattice. The type previously called diamond lattice subdivides into two types: the 'fat' diamonds and the 'thin' diamonds. The physical fact which reflects this difference is that it is not possible continuously to deform a 'fat' diamond lattice into a 'thin' diamond lattice without passing through a lattice of different type: either hexagon or parallelogram.

Note that in this case the geometric method gives no less information than the algebraic. In fact it gives much more, namely information about continuous deformations of lattices as well as about distinct types of lattices. Similar methods can be applied to the three-dimensional lattices of crystallography, and yield interesting theoretical results about deformations of crystals: I say 'theoretical' because in practice only very small deformations are possible before the crystal breaks or before the crystal structure is spoilt by a so-called dislocation.

Further reading

The remarks on differential equations in this article are deliberately brief, because they are continued in

R. L. E. Schwarzenberger, *Elementary Differential Equations* (Chapman and Hall, 1969).

For more on crystal lattices and related topics, see

Hermann Weyl, *Symmetry* (Princeton University Press, 1952).

Finally, there is a book which appeared after this article was written but which makes me understand much better the rather vague thoughts which made me write it!

Richard R. Skemp, *The Psychology of Learning Mathematics* (Penguin Books, 1971).

How Many Prime Numbers Are There?

J. H. E. COHN

Royal Holloway College, London

A prime number is defined to be an integer $p > 1$ which has only the factors 1 and p. It is a matter of common experience that such numbers exist, and there appear to be quite a number of them. Simple trial and error reveals that there are 168 primes less than 1,000. It is also well known, and not too difficult to prove that every integer exceeding 1 can be written as a product of primes, and moreover uniquely so, apart from order. With this knowledge it is easy to show that there are infinitely many primes; for if n primes are given, say $p_1, p_2, \ldots, p_n$, then the number

$$N_n = p_1 p_2 p_3 \cdots p_n + 1,$$

which exceeds all the given ones, is either itself a prime, or else has at least one prime factor p, which must clearly be different from all the given ones (why?). Thus given any set of prime numbers, it is always possible to find one more.

The proof just given is due to Euclid, but despite its great antiquity it gives rise to at least one unsolved problem. The first prime is 2. Let us take $p_1 = 2$ and actually use the above method to find some primes. We find that $N_1 = 3$ and accordingly take $p_2 = 3$. Then $N_2 = 7$ and we write (temporarily) $p_3 = 7$. (We observe that with this notation, p_3 is not actually the third prime in order of size; 5 is the third.) Then

$$N_3 = (2 \times 3 \times 7) + 1 = 43 = p_4.$$

Now

$$N_4 = (2 \times 3 \times 7 \times 43) + 1 = 1{,}807.$$

Since $1{,}807 = 13 \times 139$ we cannot write $N_4 = p_5$, and we must choose whether to call p_5 13 or 139. And in general N_n may contain even more than two distinct prime factors. So if we are to use this method to generate primes, we have to have a rule to determine which of the several prime factors is to be taken, where a choice exists. Various rules are possible, perhaps the two simplest being

(a) choose the smallest prime factor each time;

(b) choose the largest prime factor each time.

The question is whether with one of these rules, or for that matter with some other, we obtain *all* primes in this way. It is not too difficult to show that with rule (b) we do not, and in fact never obtain the prime 5. With rule (a) the question is as yet undecided, although various 'likelihood' considerations indicate that probably we do.

We have shown that the arithmetic progression (A.P.) whose mth term is $m+1$ contains infinitely many primes. What about other A.P.'s? Clearly the A.P. $6m+9$ contains no primes, since every member is divisible by 3 without equalling 3, but what about a general A.P. $am+d$ in which a and d have no common factor? The above proof with $p_1 = 2$ can quite easily be used to prove the existence of infinitely many primes in the A.P. $4m+3$. For N_n itself is of this form since $p_2, p_3, \ldots, p_n$ are all odd, and so is itself a prime of this form, or else contains a prime factor *of this form*, since otherwise it would be the product of primes all of the form $4m+1$, and so itself be of that form. Similar methods can be used for a few other special A.P.'s, but the general result that any such A.P. contains infinitely many primes is a very difficult result to prove, and is due to Dirichlet. The same question about other sequences, e.g., m^2+1, is unsolved, and quite beyond our capabilities at present, but the answer is that probably such sequences also contain infinitely many primes.

The first paragraph seems to have answered the question of the title, but in a sense rather superficially. That there are infinitely many plums (primes) in the infinite plum pudding (positive integers) may be interesting, but a potential Jack Horner would want to know more than that. How many plums are there in the first x pounds of pudding? Where is the nth plum? So we define $\pi(x)$ to denote the number of primes less than or equal to x and from now onwards denote the nth prime by p_n. Clearly $\pi(p_n) = n$ and so the growth of $\pi(x)$ with x and of p_n with n are closely related.

We have shown that $\pi(x) \to \infty$ as $x \to \infty$, and the method can actually be used to show that[1] $\pi(x) > \log\log x$ for all $x \geqslant 3$. But this result is extremely weak for it merely proves that $\pi(10^9) \geqslant 3$, whereas in fact $\pi(10^9) = 50{,}847{,}478$.

A different method yields rather better results. Consider x an integer at least equal to 3. Then

$$\left\{\prod_{p \leqslant x}\left(1-\frac{1}{p}\right)\right\}^{-1} = \prod_{p \leqslant x}\left(1-\frac{1}{p}\right)^{-1}$$

$$= \prod_{p \leqslant x}\sum_{s=0}^{\infty} p^{-s}$$

$$= \sum{}^* \frac{1}{m},$$

where the first product is taken over all primes less than x, and $\sum^*$ denotes that the summation is taken over all m which are divisible only by the primes not exceeding x. This summation accordingly certainly includes all the positive integers up to x, as well as many more, and so certainly

$$\sum{}^* \frac{1}{m} > \sum_{n=1}^{x} \frac{1}{n} > \log x.$$

Thus

$$\left\{\prod_{p \leqslant x}\left(1-\frac{1}{p}\right)\right\}^{-1} > \log x. \tag{1}$$

In particular, there must be an infinite number of primes, for otherwise as $x \to \infty$, the left hand side would eventually remain constant, whereas the right hand side

[1] All logarithms are taken to the natural base e.

tends to infinity. But more than this can be deduced. If $\pi(x) = N$ we have from (1)

$$\frac{1}{\log x} > \prod_{r=1}^{N} \left(1 - \frac{1}{p_r}\right),$$

and clearly for every r, $p_r \geqslant r+1$. Thus the rth factor in the above product exceeds $1-1/(r+1) = r/(r+1)$ and so

$$\frac{1}{\log x} > \prod_{r=1}^{N} \frac{r}{r+1} = \frac{1}{N+1},$$

whence $N > -1 + \log x$, or

$$\pi(x) > -1 + \log x, \tag{2}$$

a considerable improvement on the previous result, although still a long way from the real truth. (It is the method of deducing (2) from (1) which is responsible for this; (1) itself is quite a good result and it merely requires the $\log x$ to be doubled for the result to be true in the opposite direction.) Further improvements are possible. For example if we use instead of $p_r \geqslant r+1$ for all r, $p_r \geqslant 2r-1$ for all $r \geqslant 2$ in (1) we find that

$$\pi(x) > \tfrac{1}{8}(\log x)^2, \tag{3}$$

a better result than (2) but still numerically very poor as can be seen from taking $x = 10^9$.

The actual result is that $\pi(x)$ behaves for large x like $x/\log x$, in the sense that the ratio of the two expressions tends to 1 as $x \to \infty$. This famous theorem, the prime number theorem, was not proved until 1896 and its proof even today is rather difficult.

As mentioned above, there is a close connection between the growth of $\pi(x)$ and of p_n, and it follows from the prime number theorem that $p_n/n \log n \to 1$ as $n \to \infty$. It can then be seen that 'on average' near a number x, about $(\log x)^{-1}$ of the integers are primes, in other words 'on average' they get scarcer for larger numbers. Thus near 10^9 about 5% of the integers are primes, whereas over 16% of the first thousand integers are primes. But this talk of 'on average' hides great irregularities in the occurrence of primes. For example, the size of $p_{n+1} - p_n$ which represents the distance between consecutive primes, oscillates tremendously. On the one hand there are long runs of consecutive composite (i.e., non-prime) integers; for example the prime 370,261 is followed by no less than 111 composite integers. On the other hand, there are many pairs of primes which differ by only 2, in fact there are over 8,000 such pairs less than a million. It is conjectured that there are infinitely many such pairs, but this is a most difficult problem, and is one which is as yet unsolved.

Suggestion for further reading

G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Clarendon Press, Oxford, 1945).

On Folding a Map

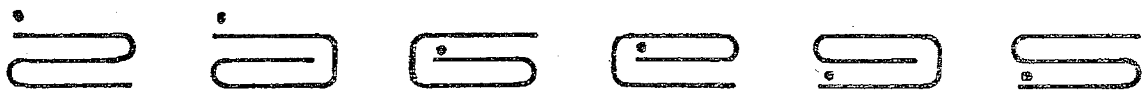
W. F. LUNNON[1]

Atlas Computer Laboratory, Chilton, Berkshire

The problem

In how many ways can you fold up a map? Our '$p \times q$-map' is an infinitesimally thin, elastic rectangular sheet, divided into pq identical leaves by $p-1$ creases one way and $q-1$ the other. A '$p \times q$-folding' is any arrangement of it such that all the leaves lie atop one another. This is achieved by folding at the creases, bending and stretching (in order to interleave), but not tearing. The 'front cover' is one side of a particular leaf (leaf 1, marked with a dot in the diagrams): it is always kept fixed in space, to ensure that we cannot get two different diagrams for the same folding. We are interested in the total number of foldings of a given $p \times q$-map, and we call this total $G(p,q)$.

Consider as an example the London Underground Railway map, for which $p = 3$ and $q = 1$. There are 6 ways to fold it, seen here from the edge:



Really, this map is only one-dimensional: we call it a p-map, $p = 3$, and talk about the total $G(p)$ of p-foldings, $G(3) = 6$. $G(p)$ is quite difficult enough—no formula for it is known—so we shall stick to it for much of this article. We shall see how to estimate it, and how to compute it by counting foldings.

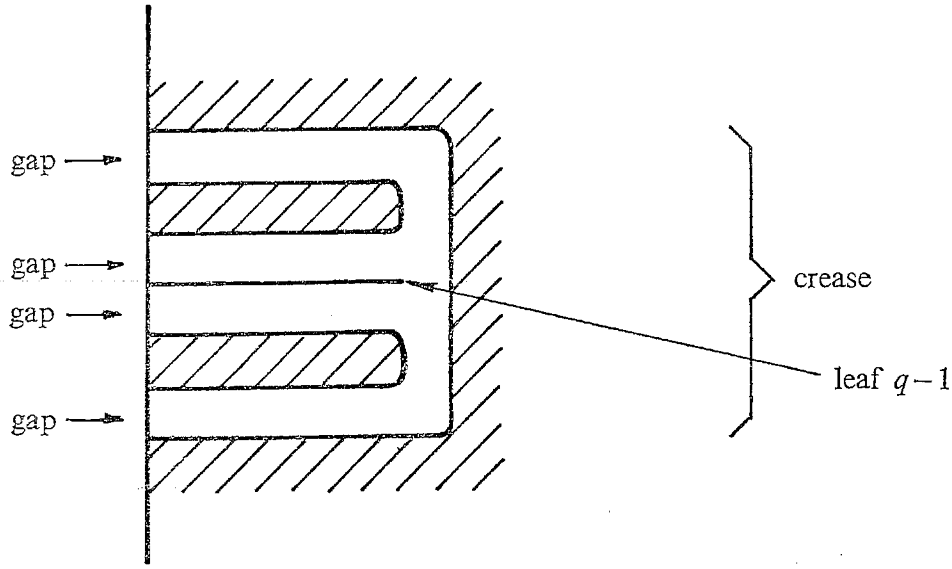
The program

To prepare map-folding for the computer we must first abstract it (since a computer can operate only on symbols) and secondly break the counting method down into simple steps (no harder than elementary arithmetic). Clearly a p-folding is completely determined by the order of its leaves. So if we number the leaves 1, 2,..., p in the natural way, an abstract description of a p-folding can be simply a list of leaf numbers from the top down: that is, a permutation of 1, 2,..., p. However, there are many ways to represent a permutation, and we shall see that the obvious way is not the best for our purpose.

We shall construct the set of all p-foldings 'by induction': that is, assuming that some genie has created for us the set of all $(q-1)$-foldings (one at a time!) we shall construct the set of all q-foldings from them; then we shall construct the single 1-folding (no great problem); then we shall apply the induction for $q = 1, 2, \ldots, p-1$. (This is called 'backtracking'.) The idea of the inductive step is simply to 'grow' a new leaf q on the tip of the leaf $q-1$, and insert this new leaf in

[1] Now at University College, Cardiff.

turn into each gap which shares the same crease. (A 'gap' is the space between two leaves adjacent in the folding; a 'crease' is the space between two leaves with a common edge.) See Figure 1, where we have chopped a folding in half and picked out this crease. In this way the 7-folding in Figure 2 grows 4 new 8-foldings.



(growing tip)

Figure 1

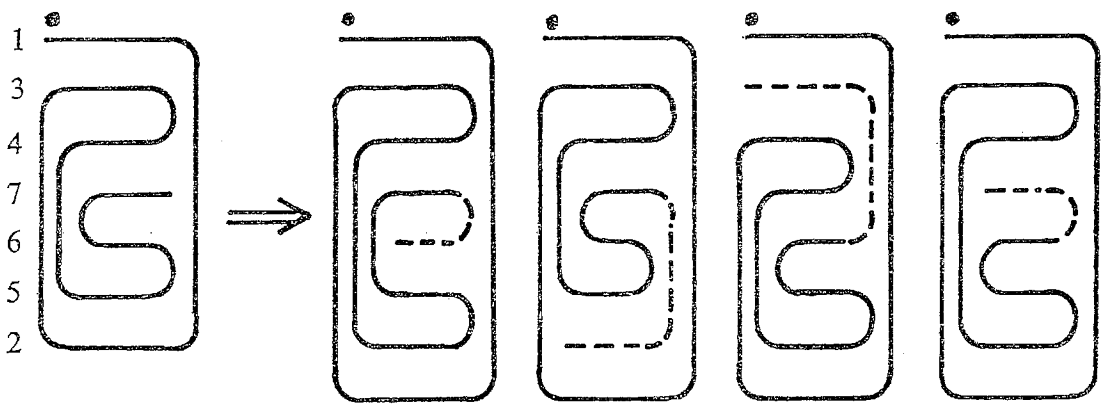


Figure 2

The computer, of course, cannot see these diagrams: like a blind beetle it scuttles about the crevices of the structure, unable to appreciate its totality. The 'inner loop' of the program—that is the part obeyed most often—is clearly going to involve getting from one gap to the next gap below in the same crease; so to speed things up, we must choose our representation of a folding to make this operation as fast as possible. We do not just keep an **array** *perm* such that *perm* [*i*] holds the leaf number of the *i*th leaf from the top. Instead we keep a pair of **arrays** *nextr* and *nextl* such that *nextr* [*i*] is the next gap below gap *i* in the same crease, on the right-hand side; and nextl does the same job on the left. Here in Table 1 are the *next* arrays for the 7-folding above; e.g., nextr (1) = 4 because the next gap on the right after the gap under leaf 1 is the gap under leaf 4.

TABLE 1

i	0	1	2	3	4	5	6	7
nextr [i]	2	4	0	3	7	1	6	5
nextl [i]	1	2	0	5	6	3	4	7

If we are inserting leaf q on the right (q even), the first gap for it will be immediately below leaf $q-1$, and the last gap immediately above. So commencing

$$j := q-1;$$

and repeating

$$j := \text{nextr}\,[j];$$

until

$$\text{nextr}\,[j] = q-1;$$

causes j to step round all the gaps for leaf q. And at each of these gaps j, performing

$$\text{nextr}\,[q] := \text{nextr}\,[j];$$

$$\text{nextr}\,[j] := q-1;$$

inserts leaf q in gap j. (We can then press on to leaf $q+1$ on the left, leaf $q+2$ on the right, ...) On return, the sequence

$$\text{nextr}\,[j] := \text{nextr}\,[q];$$

removes leaf q again. There are complications over starting leaf q off in the first gap, and seeing to the other (left-hand) side too, but we ignore them to concentrate on the basic idea.

One complication we shall discuss is leaf 0. This is invented to distinguish between the top leaf and the rest, otherwise we should have to test for it in the inner loop and waste time. Leaf 0 is above the top leaf and below the bottom leaf: it encloses the folding like a wrapper. See the examples above: e.g., *nextl* [2] = 0 and *nextl* [0] = 1, because these are the lowermost and uppermost gaps on the outside left.

A very neat and fast ALGOL program can be written along the above lines. But it can still be improved by a factor of something like a quarter of a million, in a variety of ways, most of which are rather technical: a factor of 10 by hand-coding (writing directly in the language of the particular computer); 5 by polishing the inner loops; 100 by making the program so compact that it can run as a permanent background job for very long periods; and finally 50 by counting only 'subnormal' foldings.

A 'normal' folding is one in which the front cover is at the top. In this case the first crease (between leaves 1 and 2) must turn down. If the next crease (between leaves 2 and 3) also turns down, we call the folding 'subnormal'. In the list of

3-foldings at the start of the section, the first two are normal and the first one is subnormal. It turns out that precisely one pth of p-foldings are normal and precisely one $2p$th of them are subnormal; can you prove these statements? So if we arrange for the program to construct only subnormal foldings, we shall improve its speed by a factor of $2p$. (Normal foldings are easy to guarantee: we simply miss out leaf 0 and say that the bottom leaf is immediately above the top leaf!)

A few results from our program are to be found down the left-hand column of Table 3. It has run up to $p = 24$: there are $G(24) = 258{,}360{,}586{,}368$ foldings of a map with 24 leaves.

The size of $G(p)$

Since we do not have an exact formula for $G(p)$, we ask for the next best thing: a rough estimate of it for large p. We need a theorem from analysis, called the 'Pólya–Szeged theorem'. Suppose we have some function $f(p)$, and we know that, for any p and q,

$$f(p+q) \geqslant f(p)f(q),$$

then the theorem says that as p increases the pth root of $f(p)$ approaches a limit l from below:

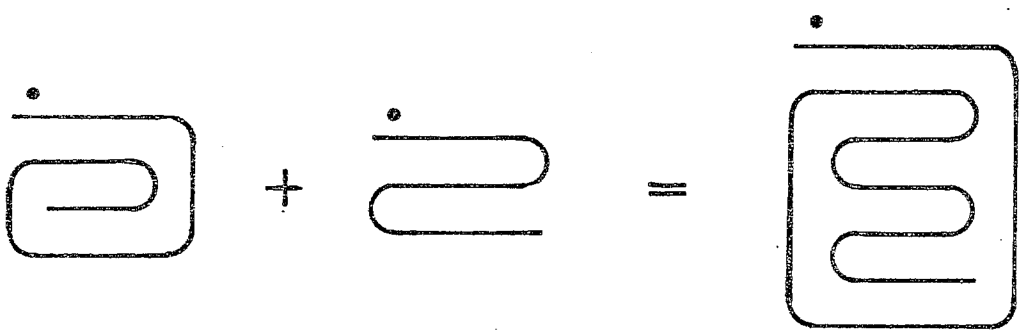
$$\sqrt[p]{f(p)} \to l \quad \text{and} \quad \sqrt[p]{f(p)} < l.$$

In other words, for large p, $f(p)$ is approximately exponential: $f(p) \doteqdot l^p$. (An example of such a function is $f(p) = 2^p$, since $2^{p+q} = 2^p 2^q$. Clearly $l = 2$ for this f.)

We cannot apply this directly to $G(p)$; but suppose we define $G'(p)$ to be the total of *normal* p-foldings,

$$G'(p) = G(p)/p.$$

Now given any pair of normal p- and q-foldings, we can make a normal $(p+q)$-folding simply by joining one on the tail end of the other: e.g., for $p = 4$, $q = 3$, we can join any normal 4- and 3- to make a 7-folding thus:



And this $(p+q)$-folding is distinct from any other so constructed. So

$$G'(p+q) \geqslant G'(p)\,G'(q),$$

and we can let $f = G'$ in the theorem to deduce that

$$G'(p) \doteqdot L^p \quad \text{for some number } L.$$

Since $G(p) = pG'(p)$, $G(p) \doteqdot L^p$ as well—the factor p is too small to interest us.

Now we know L exists, we want to find its value. Let us try to bound L from below. It is easy to see that $L \geqslant 2$: for given any p-folding we can make two distinct $(p+1)$-foldings by inserting leaf $p+1$ directly above or below leaf p. Better still we can apply the Pólya–Szeged theorem, which says that the pth root of $f(p)$ is less than l: setting $p = 24$ we find that

$$L \geqslant \sqrt[24]{10{,}765{,}024{,}432} = 2{\cdot}618\ldots.$$

Squeezing the last drop out of our numbers, we can let $f(p) = \frac{1}{2}G'(p+3)$ instead and improve this to

$$L \geqslant 2{\cdot}906\ldots.$$

Now let us try to bound L from above. For this we shall need the Catalan numbers C_n, defined as the number of ways to arrange n pairs of brackets: e.g., for $n = 3$ pairs we have $C_n = 5$ ways:

((())), (()()), (())(), ()(()), ()()().

The first few values of C_n are given in Table 2. Now given any p- and q-bracketings, we can put them together to form an $(n+1)$-bracketing $(n = p+q)$ by enclosing

TABLE 2

n	0	1	2	3	4	5	6	7	8	9	10
C_n	1	1	2	5	14	42	132	429	1,430	4,862	16,796

the p-bracketing in an extra pair of brackets, and sticking the q-bracketing on the end: e.g., for $p = q = 2$, $n = 4$ we might have

(())+()() = ((()))()().

Furthermore each $(n+1)$-bracketing can be broken down this way into just one doublet of smaller bracketings, by stripping off the first pair of brackets (arrowed). For each p and q there are just $C_p C_q$ such doublets; so

$$C_{n+1} = C_0 C_n + C_1 C_{n-1} + C_2 C_{n-2} + \ldots + C_n C_0.$$

This neat little recurrence enables us to find out all we need about C_n. We invent a function $F(x)$ whose Taylor series is C_n, that is

$$F(x) = C_0 + C_1 x + C_2 x^2 + \ldots.$$

Then

$$F(x)^2 = [C_0 C_0] + [C_0 C_1 + C_1 C_0]x + [C_0 C_2 + C_1 C_1 + C_2 C_0]x^2 + \ldots$$
$$= C_1 + C_2 x + C_3 x^2 + \ldots$$

(by the recurrence for C_n),

$$F(x)^2 = (F(x) - 1)/x;$$

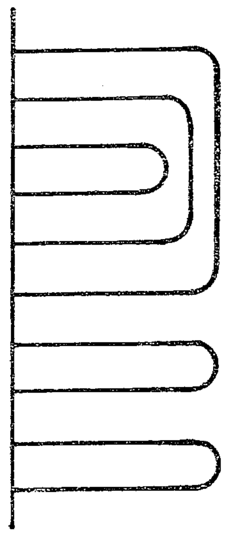
so solving for $F(x)$ we find

$$F(x) = (1-\sqrt{(1-4x)})/2x.$$

We could now expand the right-hand side into a Taylor series by the binomial theorem, and get an explicit formula for C_n—what is it? The right-hand side converges provided $|x| < \frac{1}{4}$, but not when $|x| > \frac{1}{4}$—since the value is then a complex number although the Taylor series is real. So for large n, the coefficients C_n must just outweigh x^n when $x = \frac{1}{4}$, and therefore

$$C_n \doteqdot 4^n.$$

Now to apply our knowledge of C_n to bounding L and $G(p)$. Given any folding of p leaves, we cut it in half and consider just one side. Ignoring the occasional free end, this side corresponds to a bracketing of $\frac{1}{2}p$ pairs, each pair of brackets matching a pair of leaves with a common crease:



$$\longrightarrow \quad ((()))()()$$

So there are at most $C_{\frac{1}{2}p}$ possible right-hand sides, and an equal number of left-hand sides, for a folding. Therefore

$$G(p) \leqslant (C_{\frac{1}{2}p})^2 \doteqdot 4^p.$$

That is, $L \leqslant 4$.

To sum up, we have proved that (roughly speaking)

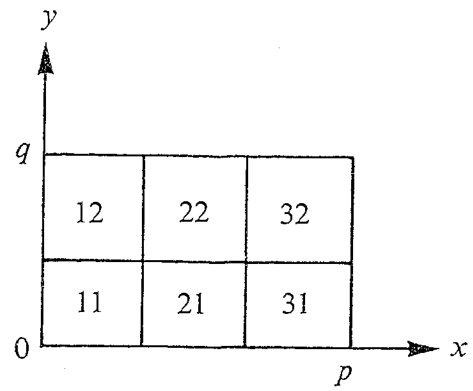
$$2{\cdot}821^p \leqslant G(p) \leqslant 4^p$$

which is not really very good. What is the value of L? Well, the ratio $G(p)/G(p-1)$ turns out very well-behaved; by assuming that it actually approaches L we can guess at the value, and this guess is

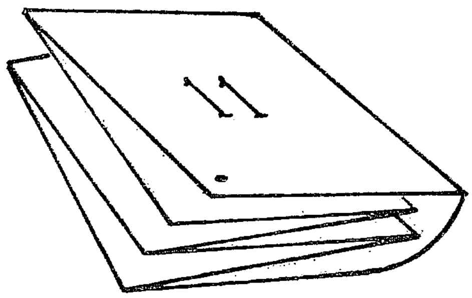
$$L = 3{\cdot}5018 \pm 0{\cdot}0001, \text{ or about } 3\tfrac{1}{2}.$$

Maps in two dimensions

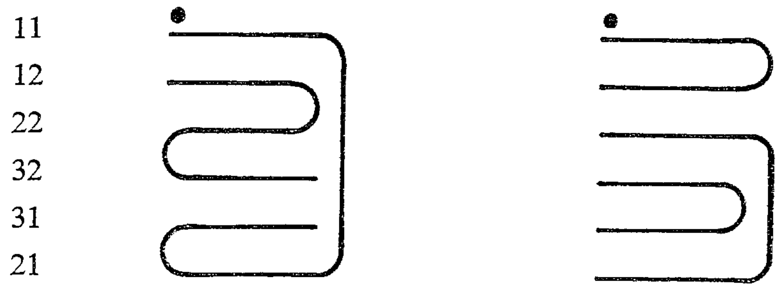
The original problem was to fold a $p \times q$-map which was properly two-dimensional, i.e., p and q might both exceed 1. For example the 3×2-map below:



Suppose we have some folding of this:



We can section it in two directions, parallel to the x- and y-axes respectively. The x-section consists of q interleaved p-foldings, and the y-section of p interleaved q-foldings:



x-section y-section

Such a pair of one-dimensional sections uniquely determines the two-dimensional folding. But more than this, *any* such pair of sections (in which the creases occur in the right places) does correspond to a genuine $p \times q$-folding. Knowing this, we can extend the one-dimensional program to the two-dimensional problem: we now simultaneously fold a pair of pq-maps, one cut into q pieces and the other into p. When growing a new leaf onto a partial folding, we insert it in just those gaps which are in the same crease in *both* sections simultaneously—there may of course fail to be any, in which case we must backtrack without having constructed any new $p \times q$-foldings.

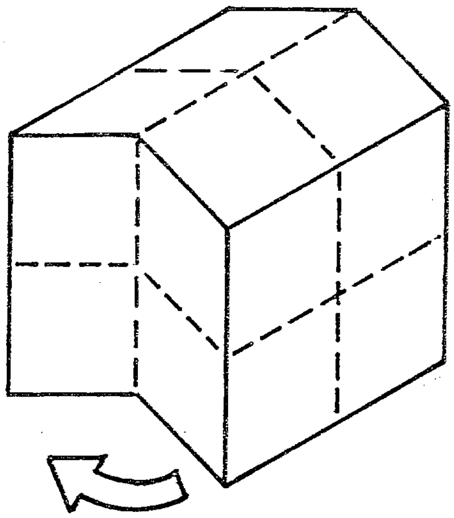
We show the results in Table 3. For the 3×2-map there are 60 foldings, and $10 = 60/3.2$ of them are normal. The numbers increase rapidly.

TABLE 3

	1	2	3	4	5
1	1				
2	2	8			
3	6	60	1,368		
4	16	320	15,552	300,608	
5	50	1,980	201,240	6,139,920	186,086,600
6	144	10,512	2,016,432	—	—
7	462	60,788	21,582,624	—	—
8	1,392	320,896	—	—	—
9	4,536	1,787,904	—	—	—
10	14,060	9,381,840	—	—	—

Maps in many dimensions

If you like the idea of higher space, try map-folding in d-dimensions. For example, a '$2 \times 2 \times 2$-map' is a $2 \times 2 \times 2$ cube whose edges are jointed at the mid-points, so that the whole thing can be 'folded' up into one octant:



As a matter of fact, it is easy to guess an answer for the $2 \times 2 \times \ldots \times 2$ map. There are d crease-planes, so you can choose the order of folding them in $d!$ ways, and fold each in two ways to get in all $2^d d!$ foldings. For $d = 1, 2, 3$ dimensions this gives 2, 8, 48; the first two are obviously correct.

As in two-dimensions, a d-dimensional folding is equivalent to a set of d one-dimensional cross-sections, and a program for $G(p,q,\ldots)$ can be written on this basis. Alas, when we run it on $2 \times 2 \times 2$-foldings it produces, not 48, but 96! Which seems a good place to end.

Suggestions for further reading

There are several references and a variety of map-folding puzzles in Martin Gardner, Mathematical games. *Scientific American* **224** (1971), 110–116.

One reference not given there is

W. F. Lunnon, Multidimensional map-folding. *Computer Journal* **14** (1971), 75–80.

Adventures in Ballistics 1915–1918. II

J. E. LITTLEWOOD

Trinity College, Cambridge

Introduction

This paper is concerned with improvements (by two orders in the errors) in the formulae available in 1915 for 'Direct Fire', i.e., for finding the range $\bar{x}(\phi)$, the 'time of flight' $\bar{t}(\phi)$, and the 'angle of descent' ω at the end, of a trajectory Γ with 'small' elevation ϕ. I use inverted commas because (a) we shall find formulae giving, e.g., $\bar{x}(\phi)$ with error $O(\phi^6)$, (b) experience shows that in practice they give results for a Γ of $\phi = 30°$ with an error of about 20 feet in 60,000. (This degree of accuracy was none-the-less a surprise. In Ballistics, 30° can be a 'small' angle.)

I arrived at the formulae shortly after, as I explained in I, I became a free-lance.

1. The air resistance is $ce^{-ky}f(v)$, where c is a constant of the gun, $f(v)$ is given by a table, and e^{-ky} is the air density at height y: it is $\frac{1}{2}$ at about 15,000 feet. Very roughly $f(v)$ is Av^2 for velocities less than v_0, that of sound, and $3Av^2$ for $v > v_0$.

2. The practically important quantities of a Γ are ϕ, the elevation, $\bar{x}(\phi)$, the range, $\bar{t}(\phi)$, the time of flight, $h(\phi)$, the greatest height, and ω, the 'angle of descent' at the point of arrival. h is needed to correct for varying air density. ω has some practical importance; also it enters our formulae. There are formulae with error $O(\phi^6)$ for $\bar{x}$, $\bar{t}$, ω, h; we will be content to discuss $\bar{x}$.

3. We will begin with some crude considerations, ignoring errors $O(\phi^2)$, which we abbreviate to $\mathscr{E}$.

Let the general point P of Γ have coordinates (x,y) (Ox horizontal), let $\tan\theta = dy/dx$ be the slope of Γ at P, and let t be the elapsed time. By choosing new units of length and time we may suppose $g = 1$, $V = 1$, where V is the muzzle velocity. By choosing a new unit for c we may further suppose that $f(V) = 1$. We have $\cos\theta = 1+\mathscr{E}$, $\tan\theta = (1+\mathscr{E})\,\theta$.

Assume a homogeneous atmosphere. Let $u = v\cos\theta$ be the horizontal velocity, whose initial value is $U = V\cos\phi = \cos\phi$. Then

$$v = (1+\mathscr{E})u, \quad f(v) = (1+\mathscr{E})f(u).$$

Resolving horizontally, we have

$$\dot{u} = -cf(v)\cos\theta = -(1+\mathscr{E})\,cf(u),$$

and so

$$\frac{du}{dx} = \frac{\dot{u}}{\dot{x}} = -(1+\mathscr{E})\frac{cf(u)}{u}. \tag{3.1}$$

Resolving normally we have

$$v\dot{\theta} = v^2/\rho = -g\cos\theta = -\cos\theta = -(1+\mathscr{E}),$$

and so

$$\left.\begin{aligned}\frac{d\theta}{dx} &= \frac{\dot{\theta}}{\dot{x}} = -(1+\mathscr{E})\frac{1}{uv} = -(1+\mathscr{E})\frac{1}{u^2},\\ \frac{d\tan\theta}{dx} &= (1+\mathscr{E})\frac{d\theta}{dx} = -(1+\mathscr{E})\frac{1}{u^2},\\ \text{and}\qquad & \\ \frac{d\tan\theta}{du} &= \frac{1+\mathscr{E}}{cuf(u)},\end{aligned}\right\} \tag{3.2}$$

by (3.1). Hence using $\int_U^u = -\int_u^U$, we have

$$\tan\theta = \tan\phi - \frac{1+\mathscr{E}}{c}\int_u^U \frac{du}{uf(u)}, \tag{3.3}$$

and so, by (3.1) (inverted)

$$y = \int_0^x \tan\theta\, dx = -\frac{1+\mathscr{E}}{c}\int_u^U \frac{\tan\theta\, u du}{f(u)},$$

whence by (3.3),

$$y = \frac{1+\mathscr{E}}{c}\int_u^U \frac{u}{f(u)}\left\{\tan\phi - \frac{1+\mathscr{E}}{c}\int_u^U \frac{du}{uf(u)}\right\} du.$$

Take $y = 0$, $u = \bar{u}$ in this:

$$\tan\phi\{S(\bar{u}) - S(U)\} = \frac{1+\mathscr{E}}{2c}\int_{\bar{u}}^U \frac{u\{J(u)-J(U)\}}{f(u)}du, \tag{3.4}$$

where

$$S(u) = \int_u^\infty \frac{udu}{f(u)}, \quad J(u) = \int_u^\infty \frac{2du}{uf(u)}. \tag{3.5}$$

Let

$$A(u) = \int_u^\infty \frac{uJ(u)\, du}{f(u)}. \tag{3.6}$$

We can write (3.4) as

$$\tan\phi\{S(\bar{u}) - S(U)\} = \{A(\bar{u}) - A(U)\} - J(U)\{S(\bar{u}) - S(U)\}. \tag{3.7}$$

Finally we have, by (3.1),

$$\bar{x} = \frac{1+\mathscr{E}}{c}\{S(\bar{u}) - S(U)\}. \tag{3.8}$$

The functions S, J, A are tabulated. By trial and error $\bar{u}$ is found from (3.7); $\bar{x}$ is then given by (3.8).

It is evident that tables of S, etc. cannot deal with a varying density. Before using them we must find an 'equivalent homogeneous density' $e^{-\rho kh}$. Finding ρ is a new problem. It was always taken for granted before 1915 that 'the average height of a Γ' is $\frac{2}{3}h$. We shall find that for $\bar{x}(\phi)$, ρ is $\rho_0 = \frac{3}{5}$ to a first approximation.

4. Siacci proposed (for homogeneous atmosphere) the formulae:

$$c\bar{x}(\phi) = S(\bar{w}) - S(W),$$
$$c\bar{x}\tan\phi = \tfrac{1}{2}\{A(\bar{w}) - A(W)\} - J(W),$$

where w is (not u, but) $v\cos\theta\sec\phi$, $W = V(=1)$. This has the advantage over the u-formulae that w is initially V, so that $f(W) = f(V)$ accurately represents the initial resistance.

We shall find that the errors in Siacci's formulae are $O(\phi^4)$.

It occurred to me to use a 'pseudo-velocity' $w = v\cos\theta/\lambda$ instead of Siacci's $v\cos\theta\sec\phi$, and if possible to find a 'best possible' λ. To secure correctness at the start the formulae from which $\bar{x}$ is found are:

$$\left.\begin{aligned} w &= v\cos\theta/\lambda, \quad W = \cos\phi/\lambda; \\ \bar{x} &= \frac{\lambda}{c}\{S(\bar{w}) - S(W)\}, \\ \tan\phi &= \frac{1}{2\lambda c}\left[\frac{A(\bar{w}) - A(W)}{S(\bar{w}) - S(W)} - J(W)\right]. \end{aligned}\right\} \tag{4.1}$$

For a Γ of elevation ϕ there is an 'equivalent' $f(v) = v^n$ ($V = 1$, $f(V) = 1$), giving the same $\bar{x}$ as the actual $f(v)$. (Consider $n = 0$ and $n = \infty$!) n will depend on ϕ; however, we aim at a λ depending as little as possible on n. We shall find $\lambda = \cos^{\frac{2}{3}}\psi$, $\psi = \phi + \frac{1}{16}n(\omega - \phi)$. The first approximation $\cos^{\frac{2}{3}}\phi$ is independent of n; in $\psi - \phi$ we have $\omega - \phi = O(\phi^2)$, and the numerical $\frac{1}{16}$ is usefully small. (In practice we used empirical n's for different types of gun.) The error in $\bar{x}$ is $O(\phi^6)$.

5. My method for λ is, first to expand $\tan\phi$ as a power series in $\bar{x}$ with error $O(\bar{x}^6)$, or, equivalently, $O(\phi^6)$; then to expand the $\tan\phi$ obtained from (4.1) as a power series in $\bar{x}$. Equating the two $\tan\phi$'s gives an equation for λ leading to $\lambda = \cos^{\frac{2}{3}}\psi$. The work is very heavy in spite of the simple final result.

It then remains to determine ρ to a second approximation for the final $\bar{x}$ to have error $O(\phi^6)$.

6. In what follows f stands for $f(v)$ and so for v^n.

The basic differential coefficients are:

$$\frac{dy}{dx} = \tan\theta, \quad \frac{d\theta}{dx} = \frac{\dot{\theta}}{\dot{x}} = -\frac{\cos\theta/v}{v\cos\theta} = -\frac{1}{v^2};$$

$$\frac{dv}{dx} = \frac{\dot{v}}{\dot{x}} = \frac{cfe^{-ky} + \sin\theta}{v}\sec\theta.$$

These enable us to expand y as a power series in x with error $O(x^7)$:

$$y = \sum_{m=1}^{6}\frac{x^m}{m!}\left(\frac{d^m y}{dx^m}\right)_0 + O(x^7).$$

The first term on the right is $x\tan\phi$. We put $y = 0$, $x = \bar{x}$, and divide by $\bar{x}$, giving a power series in $\bar{x}$ for $\tan\phi$, with error $O(\phi^6)$.

In this, certain $(d^m y/dx^m)$ have terms τ in $\sin\phi$. Now $\sin\phi = \frac{1}{2}x+\frac{1}{3}cx^2+O(x^3)$ and we modify the expansion of $\tan\phi$ by substituting this in the τ: this does not in fact affect the error $O(\phi^6)$. The final resulting formula is (dropping bars, and so $\bar{x}=x$)

$$\begin{aligned}\tan\phi &= \tfrac{1}{2}x\sec^2\phi+\tfrac{1}{3}cx^2\sec^3\phi-\tfrac{1}{12}c^2x^3(n-4)\sec^4\phi \\ &\quad +\tfrac{1}{120}x^4\{-3(n-1)\,c+(n-1)(3n-14)\,c^2+4(n-3)(n-4)\,c^3\} \\ &\quad +\tfrac{1}{360}x^5\{(n-1)(5n-40)\,c^2+2(n-3)(n-4)(3n-8)\,c^4\}-kT+O(\phi^6), \quad (6.1)\end{aligned}$$

where

$$T=\left(\tfrac{1}{12}cx^3-\frac{n-4}{20}c^2x^4\right)\sin\phi-\tfrac{1}{60}cx^4+\frac{n-4}{90}c^2x^5,$$

or, by

$$\sin\phi = \tfrac{1}{2}x+\tfrac{1}{3}cx^2+O(x^3),$$

$$T=\tfrac{1}{40}x^4c-\tfrac{5}{72}(n-4)\,x^5c^2+O(x^6). \quad (6.2)$$

I record the k-terms for a non-homogeneous atmosphere for later use in finding ρ.

7. We have next to expand $\tan\phi$ as a power series in $\bar{x}$, using the formulae (4.1): this is not heavy. Let

$$\eta = x\tan\phi-\frac{1}{2c^2}\{\Delta A-J(W)\,\Delta S\}, \quad (7.1)$$

where $\Delta(F)=F(w)-F(W)$. Using

$$\frac{dx}{dw}=\frac{-\lambda}{cw^{n-1}},\quad \frac{dw}{dx}=-\frac{c}{\lambda}w^{n-1},$$

we have

$$\frac{d\eta}{dx}=\tan\phi-\frac{1}{2c^2}\left\{\frac{J(w)}{w^{n-1}}-\frac{J(W)}{w^{n-1}}\right\}\frac{c}{\lambda}w^{n-1},$$

or

$$\frac{d\eta}{dx}-\tan\phi=-\frac{1}{2c\lambda}\{J(w)-J(W)\},$$

$$\frac{d^2\eta}{dx^2}=-\frac{1}{2c\lambda}\cdot\frac{2}{w^{n-1}}\cdot\frac{c}{\lambda}w^{n-1}=-\frac{1}{\lambda^2w^2},$$

and so on. Inserting initial values $w=W=\cos\phi/\lambda$ in the differential coefficients, we have

$$\begin{aligned}\eta &= x\tan\phi-\tfrac{1}{2}x^2\sec^2\phi-\tfrac{1}{3}x^2c\cos^{n-4}\phi\lambda^{-(n-1)} \\ &\quad +\tfrac{1}{12}x^4c^2\cos^{2n-6}\lambda^{-2(n-1)}-\tfrac{1}{36}x^5c^3(n-3)(n-4) \\ &\quad +\tfrac{1}{180}x^6c^4(n-3)(n-4)(3n-8)+O(x^7). \quad (7.2)\end{aligned}$$

$\bar{x}$ corresponds to $\eta=0$, and (7.2) with $\eta=0$, $x=\bar{x}$ gives a power series in $\bar{x}$ for $\tan\phi$ with error $O(x^6)=O(\phi^6)$. Dropping bars and comparing this with (6.5) with

$k = 0$ gives (recalling $V = 1$, $cf(V) = c$)

$$\begin{aligned}&\tfrac{1}{3}x^2 c\sec^3\phi(1-\lambda^{-(n-1)}\cos^{n-1}\phi)-\tfrac{1}{12}x(n-4)\,c(1-\lambda^{-2(n-1)}\cos^{2(n-1)}\phi)\sec^4\phi\\&\quad+\tfrac{1}{60}x^4(n-1)\,c-\tfrac{1}{24}x^4(n-1)\,c-\tfrac{1}{180}x^5(n-1)(2n-11)\,c^2\\&\quad+\tfrac{1}{120}x^5(n-1)(3n-14)\,c^2-\tfrac{1}{36}x^5(n-1)\,c^2\\&\quad=O(x^6)=O(\phi^6).\end{aligned}\tag{7.3}$$

8. Since $x = O(\phi)$ we evidently have

$$\lambda = 1+O(\phi^2).$$

Let

$$1-\mu = \cos\phi/\lambda, \quad \mu = O(\phi^2). \tag{8.1}$$

Collect like terms in (7.3) and divide by $\frac{1}{3}cx^2$; we get

$$\begin{aligned}\{1-(1-\mu)^{n-1}\}-\tfrac{1}{4}x(n-4)\,c\{1-(1-\mu)^{2(n-1)}\}+\tfrac{1}{24}x^3c(n-1)(n-6)&\\=O(x^4)\equiv O(\phi^4).&\end{aligned}\tag{8.2}$$

Since

$$1-(1-\mu)^{r(n-1)} = r(n-1)\,\mu+O(\phi^4),$$

we get from (8.2)

$$\mu\{1-\tfrac{1}{2}c(n-4)\} = \tfrac{3}{40}x^2-\tfrac{1}{24}x^3(n-6)\,c+O(\phi^4),$$

$$\mu = \tfrac{3}{40}x^2+\{\tfrac{3}{80}(n-4)-\tfrac{1}{24}(n-6)\}\,cx^3+O(\phi^4).$$

Substituting

$$x = 2\phi-\tfrac{8}{3}c\phi^2+O(\phi^3)$$

in this we find

$$\mu = \tfrac{3}{10}\phi^2-\tfrac{1}{30}cn\phi^3+O(\phi^4),$$

and so

$$\lambda = \frac{\cos\phi}{1-\mu} = 1-\tfrac{1}{5}\phi^2-\tfrac{1}{20}cn\phi^3+O(\phi^4). \tag{8.3}$$

9. In this we replace c, which overemphasises the initial behaviour, by something depending on the Γ as a whole. We have easily, by differentiating $\tan\theta$ twice with respect to x,

$$\omega-\phi = \tfrac{1}{3}cx^2+O(\phi^3) = \tfrac{4}{3}c\phi^2+O(\phi^3). \tag{9.1}$$

Substituting for c from (9.1) in (8.3) gives

$$\lambda = 1-\tfrac{1}{5}\phi^2-\tfrac{1}{40}n\phi(\omega-\phi)+O(\phi^4),$$

$$\lambda = \cos^{\frac{2}{5}}\psi+O(\phi^4), \quad \psi = \phi+\tfrac{1}{16}n(\omega-\phi)$$

It is now easily seen that with $\lambda = \cos^{\frac{2}{5}}\psi$ the error in x given by (4.1) is $O(\phi^6)$.

A similar analysis applied to Siacci's formulae shows that the error in x is $O(\phi^4)$, and no better.

10. So far we have necessarily assumed a homogeneous atmosphere. There remains the problem of ρ such that (4.1) with $\lambda = \cos^{\frac{2}{5}}\psi$ and $c_1 = ce^{-\rho kh}$ for c, give $\bar{x}$ with error $O(\phi^6)$. The necessary materials are to hand in (6.1) and (6.2). We equate (6.1) with c to (6.1) with c_1 for c and $k = 0$. Straightforward calculation, using $x = 2\phi - \frac{8}{3}c\phi^2 + O(\phi^3)$, gives

$$\frac{c_1}{ck} = 1 - \tfrac{3}{10}\phi^2 + \tfrac{1}{30}(11n-12)\,\phi^3 c + O(\phi^4). \tag{10.1}$$

The maximum height h is the maximum of y for varying t. From $\dot{y}_0 = \sin\phi$, $\ddot{y}_0 = -(c\sin\phi + 1)$ we have

$$y = t\phi - \tfrac{1}{2}t^2(c\phi+1) + O(t^3),$$

and so

$$h = \tfrac{1}{2}(\phi^2 - c\phi^3) + O(\phi^4).$$

From this, (10.1) with $c_1/c = e^{-\rho kh}$, and (9.1) we find

$$\rho = \rho_0 + (a - bn)(\omega - \phi)/\phi.$$

This and $\lambda = \cos^{\frac{2}{5}}\psi$ give $\bar{x}(\phi)$ with error $O(\phi^6)$, where $\rho_0 = \frac{3}{5}$, and a and b are numerical constants whose exact values need not concern us.

11. I will finally say something about formulae for $\bar{t}$ and ω (for which however I have only the first approximation ρ_0 for ρ). For t we need a further Ballistic function

$$T(w) = \int_w^\infty \frac{dw}{f(w)}.$$

We now have the scheme:

$$w = v\cos\theta/\lambda, \quad W = \cos\phi/\lambda; \quad \lambda = \cos^m\psi;$$

$$\bar{x}(\phi) = \frac{\lambda}{c}\{S(w) - S(W)\}, \tag{11.1}$$

$$\bar{t}(\phi) = \frac{1}{c}\{T(w) - T(W)\}, \tag{11.2}$$

$$\tan\phi = \frac{1}{2\lambda c}\left[\frac{A(w)-A(W)}{S(w)-S(W)} - J(W)\right], \tag{11.3}$$

$$-\tan\omega = \tan\phi - \frac{1}{2c\lambda}\{J(w) - J(W)\}. \tag{11.4}$$

$\bar{x}$, $\bar{t}$, ω are obtained from (11.3), and (11.1), (11.2) and (11.4) respectively. The λ and ρ_0 are:

for $\bar{x}$, $m = \frac{2}{5}$, $\psi = \phi + \frac{1}{16}n(\omega - \phi)$, $\rho_0 = \frac{3}{5}$;
for $\bar{t}$, $m = \frac{3}{5}$, $\psi = \phi + \frac{1}{16}(n-1)(\omega - \phi)$, $\rho_0 = \frac{2}{5}$;
for ω, $m = \frac{1}{5}$, $\psi = \phi - \frac{1}{4}(n-2)(\omega - \phi)$, $\rho_0 = \frac{4}{5}$.

In none of these cases is $\rho_0 = \frac{2}{3}$! In all cases $m+\rho = 1$, a fact for which I have no explanation.

I may mention a very paradoxical case of a 'ρ_0'. For fire up an inclined plane of slope S, with small tangent elevation E (i.e., $\phi - S$) the 'obvious' mean value of y is $\frac{1}{2}H$, where H is the height of the point of arrival. None-the-less in finding t given E we have $\rho_0 = 0$. See the section on Ballistics in my book *A Mathematician's Miscellany*, or a more complete account in *Mathematical Gazette*, **52** (1968), 132–134.

Corrections to Part I (*Vol.* 4, *pages* 31–38)

Page 33, line 4: for η read $\dot{\eta}$.
Page 33, line 6 (equation (4.3)) and also line 12: delete $\int_0^t$ and dt.
Page 35, delete lines 9 and 10.
Page 35, line 11: should begin "What we do in method 1 is . . .".
Page 35, line -7 (equation (10.2)): the second X should read X_1.

Problems and Solutions

Readers who have not yet reached the age of 20 on 1 April 1972 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

4.5. A thin rigid rod is carried along a straight corridor of width u. At the end of the corridor is one at right angles to it of width v, and it is required to carry the rod round the corner. If the rod must always be held in a horizontal position, what is the maximum possible length of the rod?

4.6. If x, y, z denote the lengths of the sides of a triangle, show that

$$3(yz+zx+xy) \leqslant (x+y+z)^2 < 4(yz+zx+xy).$$

4.7. Let A be a square matrix with real elements. Let s be a real number greater than or equal to all the row and column sums. Show that each element of the matrix can be replaced by one at least as large so that all the row and column sums are s.

(*Note*: The 2×2 matrix $\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}$ has row sums 3, 7 and column sums 4, 6.)

4.8. Let n be a positive integer and let $a_1, a_2, \ldots, a_n$ be any real numbers greater than or equal to 1. Show that

$$(1+a_1)(1+a_2)\ldots(1+a_n) \geqslant \frac{2^n}{n+1}(1+a_1+a_2+\ldots+a_n).$$

Solutions to Problems in Volume 4, Number 1

4.1. Show that $\min\{m^{1/n}, n^{1/m}\} \leqslant 3^{1/3}$ for all positive integers m, n.

Solution by A. H. Rodgers (*Gonville and Caius College, Cambridge*)

If $k \geqslant 3$, then

$$\frac{(k+1)^3}{k^3} = \left(1+\frac{1}{k}\right)^3 \leqslant \left(\frac{4}{3}\right)^3 = \frac{64}{27} < 3,$$

whence by induction $n^3 \leqslant 3^n$ for all positive integers n, because the result is obvious when $n \leqslant 3$. We can suppose that $m \leqslant n$, in which case

$$\min\{m^{1/n}, n^{1/m}\} \leqslant m^{1/n} \leqslant n^{1/n} \leqslant 3^{1/3},$$

by the above.

Also solved by E. May (University of Sheffield), C. Goldthorpe (Leeds Grammar School), J. E. Macey (Nottingham High School), J. R. Partington (Gresham's School, Holt), A. H. Muhr (Grammar School for Boys, Cambridge), A. David (University of Surrey), N. I. Shepherd-Barron, The College, Winchester.

4.2. Teams $T_1, T_2, \ldots, T_n$ take part in a tournament in which every team plays every other team just once. One point is awarded for each win, and it is assumed that there are no draws. Denoting by $s_1, s_2, \ldots, s_n$ the total scores of $T_1, T_2, \ldots, T_n$ respectively, show that, for $1 \leqslant k \leqslant n$,

$$s_1 + s_2 + \ldots + s_k \leqslant nk - \tfrac{1}{2}k(k+1).$$

Solution by J. E. Macey (Nottingham High School)

$T_1, T_2, \ldots, T_k$ play $\frac{1}{2}k(k-1)$ games amongst themselves and $k(n-k)$ games against $T_{k+1}, T_{k+2}, \ldots, T_n$. Now $s_1 + s_2 + \ldots + s_k$ is greatest when all the latter games are won by $T_1, T_2, \ldots, T_k$, so that

$$\begin{aligned} s_1 + s_2 + \ldots + s_k &\leqslant \tfrac{1}{2}k(k-1) + k(n-k) \\ &= nk - \tfrac{1}{2}k(k+1). \end{aligned}$$

Also solved by C. Goldthorpe (Leeds Grammar School), A. H. Rodgers (Gonville and Caius College, Cambridge), P. B. Quigley (Ampleforth College, York), D. C. Whitley (Portora Royal School, Enniskillen), M. B. Cashman (St Cuthbert's Grammar School, Newcastle), J. A. Blake (Rugby School), A. David (University of Surrey), N. I. Shepherd-Barron, The College, Winchester.

4.3. A family has n children, where $n > 1$. Let A be the event that the family has at most one girl and B the event that not every child is of the same sex. Determine the value of n for which A and B are independent events.

Solution by J. E. Macey (Nottingham High School)

Let $p(X)$ denote the probability that the event X occurs. Then

$$p(\mathrm{A}) = \frac{1+n}{2^n}, \quad p(\mathrm{B}) = 1 - \frac{2}{2^n}, \quad p(\mathrm{A} \cap \mathrm{B}) = \frac{n}{2^n}.$$

The events A and B are independent if and only if

$$p(A \cap B) = p(A)\,p(B),$$

which is equivalent to

$$2^{n-1} - n - 1 = 0.$$

The only positive integer n which satisfies this is $n = 3$.

Also solved by C. Goldthorpe (Leeds Grammar School), A. H. Rodgers (Gonville and Caius College, Cambridge), J. R. Partington (Gresham's School, Holt), N. I. Shepherd-Barron, The College, Winchester.

4.4. A train leaves the station punctually. After having travelled for 8 miles, the driver looks at his watch and sees that the hour hand is directly over the minute hand. The average speed over the 8 miles is 33 miles per hour. At what time did the train leave the station?

Solution by A. H. Rodgers (Gonville and Caius College, Cambridge)

Let the time that the driver looks at his watch be m minutes past h $(0 \leqslant h \leqslant 11)$. Since the minute hand makes 12 revolutions for 1 revolution of the hour hand, we have

$$12(m - 5h) = m,$$

so that

$$m = \tfrac{60}{11}h.$$

But the journey took $14\frac{6}{11}$ minutes, and the train left the station at a whole number of minutes past the hour. Hence $m-\frac{6}{11}$ must be an integer, i.e.

$$\begin{aligned} \tfrac{60}{11}h-\tfrac{6}{11} &\quad \text{is an integer,} \\ 5h \equiv 6 &\quad (\text{mod } 11), \\ 5(h-10) \equiv 0 &\quad (\text{mod } 11), \\ h-10 \equiv 0 &\quad (\text{mod } 11), \\ h = 10 &\quad (\text{since } 0 \leqslant h \leqslant 11). \end{aligned}$$

Thus $m = 54\frac{6}{11}$, and the train left the station at 10.40 (a.m. or p.m.).

Also solved by P. T. Harrison (Forest Hill School, London), J. E. Macey (Nottingham High School, P. B. Quigley (Ampleforth College, York), J. R. Partington (Gresham's School, Holt), D. C. Hogg (Grammar School for Boys, Cambridge), D. C. Whitley (Portora Royal School, Enniskillen), M. B. Cashman (St Cuthbert's Grammar School, Newcastle), Janet Marshall (Monkwearmouth School, Sunderland), G. N. Sarma (Mrs A. V. N. College, Visakhapatnam, India).

The Teaching of Probability and Statistics. Edited by LENNART RÅDE. Almqvist and Wiksell, Stockholm, 1970. Pp. 373. About £7·50.

This is the report of an international conference on the teaching of Probability and Statistics in schools, held at Carbondale, Illinois, in the U.S.A. in 1969 under the auspices of the Comprehensive School Mathematics Program.

There are 18 sections dealing with various aspects of the subject; from the details given, all the authors are university mathematicians with one exception. It is significant (perhaps even statistically so) that this one contribution by Arthur Engel of W. Germany contains in its 64 pages the most practical part of the book for a school teacher. In this section there is a mine of examples which a teacher may dig at for classroom use at all levels of the secondary school.

The rest of the book is not without usefulness, but things are rather more thinly spread. Clearly a work of this type will have omissions, duplications and inconsistencies. The emphasis is more on probability than statistics, and the levels of treatment bewildering. There are too many syllabuses quoted from different countries, one or two would have been enough.

There is an interesting chapter on non-parametric statistics (sadly neglected in schools) which could have been carried further, and some ideas for statistical experiments given by Hilda Davies (Sheffield). The Dutch contribution on the aims of teaching probability is interesting, and J. Gani (Sheffield) has one on mathematical model building with some interesting examples.

The American contributions are sometimes irrelevant to teachers in England because of the emphasis on formal algebra, and the assumption that no calculus has been done. But here too one can find plenty of good ideas such as the surprise function $S(p)$ where p is the probability of an event. The axioms are:

(1) $S(1) = 0$.

(2) If $x > y$, then $S(x) < S(y)$.

(3) S is a continuous function.

(4) $S(xy) = S(x) + S(y)$.

The last gives the game away, but read the book if you want to find out more!

D. V. Lindley (London) has a chapter on 'A Non-Frequentist View', in which the most memorable sentence is 'There is no such thing as probability, only conditional probability'. Professor Lindley's remarks on p. 248 on frequency probability spaces (fps) are thought-provoking. A thumbtack is spun and lands point up 7 times out of 16. What is the probability of its landing point up on a further throw? The answer depends on the fps; if the 16 is fixed it is 7/16, if the 7 is fixed it is 6/15.

Perhaps the most intriguing piece of information in the book is that left-handed coconut trees bear more fruit than right-handed ones. It is a pity that more examples in statistics do not have an $S(p)$ as high as this!

Once the initial disappointment that these contributions will tell one what to do and not to do has worn off, the book has a lot to offer. But there is much that it gives no help on; for example, on sampling and teaching of standard deviation, it is silent.

The book is well produced, though not without errors. The most serious I spotted is that the titles of the diagrams on pp. 244–245 have become reversed, somewhat spoiling an excellent example. There is a really extensive bibliography, but why, oh why, do publishers produce books like this without an index?

University of Nottingham K. E. SELKIRK

Advanced General Statistics. By B. C. ERRICKER. English Universities Press, London, 1971. Pp. viii+350. £1·85, card covers.

This book is intended to fill a notable gap in the literature, namely to provide for students with O-level (or a bit more) in mathematics, 'a book on statistics that they could read on their own and would give them a good working knowledge of the subject'. Although it is attractively produced, written in a terse, no nonsense style, and covers the ground for all existing A-level syllabuses (except for continuous probability models, which get a page and a half), the book has three major failings, and I would not wish to use it.

The most important fault is that the author consistently devotes too little time to the question of what model to use; it is only in examination questions that the answer to this is immediately obvious. In particular, he does not distinguish between a statistical population and a probability model, and only glances at the parameters of the continuous case for the Normal curve (misleadingly spelt with a small *n*), and for three simple cases inserted without explanation or application at the end of the Normal chapter.

Secondly, the book contains definite mistakes. The crucial point that a histogram is essentially a frequency *density* diagram is not appreciated, so that (p. 13) the author is wrong in labelling the vertical axis 'Class Frequency', when it is in fact the frequency per five-year class; at this level it would be better divided by five and made into a frequency density. Nor is he helped by the fact that the text says '0·8 cm. for 10 years', but the diagram reduces the interval to 0·73 cm. The multiplication law omits the important word (and idea) 'independent', and as quoted (p. 93) is wrong. The answer to a worked example about dealing yourself a complete suit is wrong (p. 109). The χ^2 curve on 2 d.f. (simply the negative exponential, of course), and others, are drawn wrong (pp. 188, 287). It is at least misleading to say that a hypothesis not significant at the 10% level is 'very probably true' (p. 190). The 'too good to be true' treatment is also defective—'the hypothesis was manufactured to fit the data' is normal good practice (p. 191), though I know what he means! There is a suggestion that the χ^2 test decides which of two hypotheses fits data better (p. 73). He also uses 'standardized curve' for one that is not (p. 235), and 'normal variate' for $m_1 - m_2$ (p. 248), and among other infelicities 'calculate' for both 'estimate' and 'investigate'; there is confusion between a set and its size (though '—' is not an easy set operation at the best of times) (p. 97).

Thirdly, though this is a very personal matter as his pupils may be very different from mine, I find the author's pace uneven. He introduces difficult words and ideas without a tremor, and readily launches into pages of calculation, as though concepts were easy to acquire but algebra difficult to follow through. Among other things I should like to see altered (apart from misprints in the text—I have not checked many answers) are pages of easily avoidable heavy algebra. He has introduced generators, so why does he not use $G'(1)$ and $G''(1)$ to simplify the calculation of Binomial and Poisson parameters (pp. 128, 149)? He introduces the idea of standardization, so why does he not use the similar idea of transforming a population by $t = ax+b$ to simplify his treatment, *inter alia*, of $s_{xy}/s_x s_y$ (pp. 7, 81)? Must he complicate the first look at $\sigma/\sqrt{n}$ by sampling without replacement—or equivalently by using a finite population (p. 217)? Why in the linear case does he not find the regression line by taking the origin at the mean and proceeding intuitively by minimizing a simple quadratic, instead of solving two simultaneous equations produced out of a hat (p. 51)? Why are so many of his non-examination questions ten years old?

In short, this book is an attempt to fill a serious gap: good in itself, but disappointing overall. The author would have done better to take into account some of the recent developments in teaching the subject (such as the relevant chapters in *SMP Advanced Mathematics*, Books 3 and 4, for all their faults), and to persuade someone more sensitive to the use of language and ideas to look at the text before publication.

Marlborough College L. E. ELLIS

Let's Look at the Figures—The Quantitative Approach to Human Affairs. By D. J. BARTHOLOMEW and E. E. BASSETT. Penguin Books, Harmondsworth, Middlesex, 1971. Pp. 319. £0·40.

This is a delightful book. It will be of interest to students of human affairs who wish to discover how mathematics can be applied to politics, education and opinion polls, among other topics. It should prove equally fascinating to mathematics students who want to become familiar with the uses of mathematics and statistics in the social sciences.

In their introduction, the authors stress the importance of mathematical models. Part I of the book is concerned with how to measure quantities such as swings in elections, the uncertainty of events, and human ability. These are studied within the framework provided by mathematical models. Political behaviour, uncertainty and the notion of probability, variability as an introduction to elementary statistics, human variation, the measurement of I.Q. and its statistical distribution are discussed in the first four chapters.

Part II is concerned with models, and predictions which can be made from them. A quantitative approach to elections is sketched, a theory of committees outlined, the statistics of wars considered (and such problems raised as whether the outbreaks of war per year follow a Poisson distribution), and the games theory approach to conflict illustrated.

Part III deals with sampling and inference. Opinion polls, sampling methods, measures of association in both a population and a sample are considered.

Part IV studies random processes. Some processes are such that the data arising from them take the form 0 or 1, as for example in the recording of a dry or rainy day; others are measured on a continuous scale, like the daily mean temperature in a city, recorded over an annual period. In a final chapter, the authors review the salient points of the book and end with the conclusion that 'the quantitative approach is one which complements but does not replace more traditional methods of exploring the richness and diversity of human experience. It has its limitations but in our time it is a tool which we can ill afford to neglect'. How very right the authors are can be judged by anyone who reads this book. Strongly recommended.

University of Sheffield J. GANI

Conceptual Models in Mathematics. By K. E. HIRST and F. RHODES. George Allen and Unwin Ltd., London, 1971. Pp. 182. £3·00 (cloth), £1·65 (paper).

This is No. 5 in the publisher's series of Mathematical Studies, and is derived from a first-year series of lectures given at Southampton University. Perhaps the most useful way to review it will be to summarise the contents.

The first chapter is an excellent, if rather lengthy, survey of the 'Language of Sets', and covers what a good O-level candidate would know. The second chapter is on 'Mathematical Proof' and is a most illuminating survey of logic, and in particular of tautology.

So far, so good; but in the next chapter 'Relations and Functions' things get harder. The sections deal with ordered pairs and cartesian products, relations, equivalence relations, partial order relations, functions, special functions and finite and infinite sets. My guess is that many sixth formers (and teachers too) will founder somewhere here, lost in a morass of notation and language and missing the firmer ground of concrete examples. I can only suggest they hang on, or come back for a second try, for it is a worth while chapter to have a grasp of.

Chapter 4, on 'Mathematical Theories', deals with binary operations, groups, metric spaces and measure spaces and is not, I think, any harder. Chapter 5 allows one to take breath and deals with 'Probability Theory'. Most of this could have been omitted and the extra space given to expanding Chapters 3 and 4. As it is, the book finishes almost with a whimper, on a topic it cannot hope to cover fully.

Nevertheless, this is a book to have around and to give sixth formers to read. It ties up a lot of ideas which are often scattered in various places, and on the whole it refuses to be sidetracked.

The exercises are useful, but it is a pity that a number of answers, particularly in the first two chapters, are incorrect; this is bound to put off readers who lack confidence in their own understanding. It is also instructive to notice that the paperback edition of the first book in this series cost less than half the price in 1966. Perhaps an extra section on inflation functions should have been included! The school library ought to be able to afford a copy, even if it is beyond the individual's pocket.

University of Nottingham K. E. SELKIRK

Sets and Symbolic Logic and **Relations and Functions.** By H. T. COMBE. Studies in Mathematics Series, Ginn & Co. Ltd., London, 1970. Pp. 138 and 183. £0·68 and £0·80.

The author is head of the mathematics department at Stranmillis College, Belfast, and these two little books would appear to be aimed in the first instance at student teachers. The first contains chapters on definition, notation and the usual operations with sets; numbers—naturals, integers and rationals; on combinatorics; symbolic logic and valid argument; and Boolean Algebra. There is a fair number of exercises and some interesting topics for investigation. I found the balance of the material somewhat unusual; the collection in a single book of a number of different areas of mathematics in which sets are important has meant that some of the topics are not developed to any great depth. For example, the integers and rationals are introduced via pairs of naturals with defined addition and multiplication, but the truth of the Principle of Induction is stated to be 'evident'. Later in the book, it seems a pity to introduce the laws of Boolean Algebra without then using them very fully in the solution of logical or circuit problems.

The second of these volumes covers general relations and functions; linear functions (with linear programming); polynomials; derivative and integral; logarithmic and circular functions. This provides quite a useful survey, though the early part seems rather overweighted with theory as compared with applicable ideas. Again there is a useful collection of exercises, but some lack of consistency in the treatment, although this may be a subjective impression. A good up-to-date definition for the circular functions lapses into a cumbersome treatment of their addition formulae, and there is no comment on the justification for writing e^x with the x as an index. There are many, however, who will find in both of these books a useful collection of material in a compact form.

Clifton College of Education, Nottingham A. W. BELL

Probability. By T. V. M. MCKIRGAN. Studies in Mathematics Series, Ginn & Co. Ltd., London, 1970. Pp. v+138. £0·68.

This book provides a clear and reasonably concise introduction to the more elementary parts of probability. The author expresses a hope that the use of set theory will lead to a much clearer understanding of the subject—and the book gives grounds that this hope might be fulfilled. Certainly the subject matter provides scope for the application of set theory at an early stage, although the problems are limited to solutions by Venn–Euler diagrams.

There are sufficient worked examples and simple exercises throughout the text to enable the average sixth former to use the book with a minimum of supervision.

Answers are provided to the quite ample exercises at the ends of the chapters. The Binomial, Poisson and Normal distributions are considered fairly fully, and the book ends with a brief introduction to Markov chains and stochastic matrices.

At times the distinction between topics which are deemed to be within the scope of the book (Probability) and those which are not (Statistics) becomes artificial. The author's reluctance to trespass upon the preserves of his companion volume in this series—*Statistics*—is understandable, but not always helpful to the reader. It is unfortunate that $\bar{S}$ is used to denote 'not S' when considering branch diagrams while a few pages later $\bar{X}$ denotes the mean of X. The mathematically inclined reader will not enjoy such remarks as: 'As 0! does not mean anything, we define it . . .', or again abruptly 'But $\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}z^2\right] dz = \surd(2\pi)$', this latter with no reference to a reputable mathematical hatter, but such lapses are not too frequent.

This is a suitable text for sixth form use; it also deserves a place on the library shelf, where it would be useful either as a revision text or as a topic book for pupils following a traditional A-level course.

University of Nottingham M. J. BARBER

Sampling Inspection. By M. BRUCKHEIMER and A. STEWARD. Chatto and Windus, London, 1971. Pp. vii + 63. £0·60.

This booklet, the second in a series on Statistical Topics, is divided into three sections. The first is the main text and uses a hypothetical but not unrealistic problem to develop ideas about combining probabilities, and the Binomial distribution. At suitable points in the text references are inserted to some 40 exercises which form the second section. The third section consists of detailed answers to these exercises.

The initial commercial problem is sustained through the 25 pages of text, with an occasional appeal to a pack of cards for additional examples. The exercises are often used conventionally to illustrate the text, but some of them introduce new topics allied to ideas in the text. The sixth form reader may need some guidance to help him to distinguish between the fundamental mathematical ideas and the less important commercial wrappings. With this proviso, the booklet will be a useful addition to a fifth or sixth form library.

University of Nottingham M. J. BARBER

Editorial comment. It would seem that the Reviews Editor was mistaken in offering the book below for review. However, the reviewer makes an important criticism in the last paragraph which, appropriately modified, could be levelled at some current trends in the teaching of mathematics and science. The review is offered, therefore, as a starting point for discussion.

J. V. ARMITAGE

Physics Through Experiment. By RICHARD BAMBERGER. Sterling Publishing Co. Inc., New York, 1969. The Oak Tree Press Ltd., distributed in the U.K. by Ward Lock Ltd. Pp. 160. £1·25.

This is a very elementary book. It starts with basic measurements of length and volume, and progresses, with brief doses of information and instruction and an abundance of simple experiments, to include time, matter, heat, weather, noise and music, and magnetism.

It does not match any known syllabus, nor is it meant to. It really offers an experimental dabble in interesting branches of Physics. For a young child, deprived of good schooling, who is happy to work through a course of simple experiments at home, the book has obvious merits. It is never long-winded; the illustrations are good; the experiments are all simple and clearly described.

It is unlikely to be of much use to an older pupil. Would he really need to make lots of small cubes to fit into a big cube to understand volume? Would he be happy to construct a makeshift graduated cylinder from a tumbler, a balance from a ruler, or a spirit level from a medicine bottle? He would surely be familiar enough with the real thing to know how it works.

Although, with careful selection, this book might help the domestic dabbler, it is unsuitable for the serious student. There is not the reinforcement material in it. It is true the experiments give glimpses of the principles behind them, but they are not backed up. There are no concise summaries of the steps passed; there are no numerical exercises. Questions are asked occasionally to stimulate thought, but there is no way of telling if the answers are correct.

The publishers claim that this book is written for the person who wants to teach himself. Surely such a book should not let the student just perceive knowledge: it should then rub the knowledge in. This book may allow its follower to obtain many quick glimpses of truths through experiments, but because the images are not fixed by reinforcement, they would fade into obscurity. Only the interest would remain.

Shrewsbury School, The Schools, Shrewsbury RODNEY HOARE

Notes on Contributors

Paul Jackson has been a Lecturer in Statistics at the University College of Wales, Aberystwyth, since 1967. Prior to that he spent ten years in India, as a Lecturer in the Mathematics and Statistics Department of St John's College, Agra. In 1969–70 he was a Visiting Research Fellow with Educational Testing Service, Princeton, New Jersey, and he now acts as a Statistical Consultant to the American College Testing Program in Iowa City.

D. W. Sharpe, a graduate of Oxford and Sheffield, is a Lecturer in Pure Mathematics in the University of Sheffield. He has also taught in a summer school organized by the University of Illinois. David Sharpe is an algebraist, and joint author of a book about to be published in the series of Cambridge Tracts.

R. L. E. Schwarzenberger is a graduate of Cambridge. He has taught in the University of Liverpool and is now a Professor of Mathematics in the University of Warwick. His research centres on geometry and topology, and his interest in mathematical education is reflected in his book on differential equations referred to at the end of his article.

J. H. E. Cohn has taught at Bedford College and is now a Reader in Mathematics at Royal Holloway College, University of London. His many research papers deal primarily with number theory, inequalities, and differential equations. He has repeatedly taken part in the annual sixth form conference held at Bedford College, reports of which should be of great interest to our readers. See *Exploring University Mathematics* (ed. N. J. Hardiman, Pergamon Press), Vol. 1 (1967), Vol. 2 (1968), Vol. 3 (1969).

W. F. Lunnon lectures in computing at University College, Cardiff. His principal amusement is to expend appalling amounts of computer time on problems like map-folding. His hobby is bicycle riding; he is married with three children.

J. E. Littlewood, Fellow and Copley medallist of The Royal Society, De Morgan medallist of the London Mathematical Society, honorary doctor or member of many universities and academies, is the outstanding mathematical analyst of his generation. Born in 1885, he has been a Fellow of Trinity College, Cambridge, since 1908 and Rouse Ball Professor of Mathematics from 1928 to 1950. Littlewood's papers in analysis and number theory, of which over a hundred were written in collaboration with the late G. H. Hardy, have a striking power which to mere mortals seems nothing short of miraculous. Now, at the age of 86, Littlewood retains unimpaired his zest for mathematics; and the quality of his delight in mathematical activity may be gleaned from his charming booklet *A Mathematician's Miscellany* (Methuen, London, 1953).

Contents

PRICES (*postage included*)

	Individual annual subscription (1 volume of 2 issues)	Bulk annual subscriptions (5 or more copies of 1 volume of 2 issues)	Single issues
U.K. only	£0·40	£0·35	£0·23
Australia	$A1·45	$A1·25	$A0·80
U.S.A.	$1·60	$1·40	$0·90

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

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

or to Oxford University Press Branches Overseas.

Printed in England by John Wright & Sons Ltd, Bristol