GREAT IDEAS OF MODERN MATHEMATICS: THEIR NATURE AND USE

JAGJIT SINGH

GREAT IDEAS OF
MODERN MATHEMATICS

Their Nature and Use

By

JAGJIT SINGH

DOVER PUBLICATIONS, INC.
NEW YORK

Great Ideas of Modern Mathematics: Their Nature and Use is a new work, first published by Dover Publications, Inc., in 1959. It is published simultaneously in the United Kingdom by Hutchinson & Company, Ltd., under the title *Mathematical Ideas —Their Nature and Use.*

International Standard Book Number: 0-486-20587-8
Library of Congress Catalog Card Number: 60-1473

Manufactured in the United States of America
Dover Publications, Inc.
180 Varick Street
New York, N.Y. 10014

CONTENTS

ACKNOWLEDGMENTS

In writing a book of this kind one is inevitably indebted to more people than one knows. The authors that one has to read to provide the background material are too numerous to be listed. Nevertheless, I cannot forgo the pleasure of acknowledging with gratitude my debt to Professor Kenneth O. May, Sir K. S. Krishnan, F.R.S., Prof. Mahalanobis, F.R.S., Prof. R. Vaidyanathanswamy, Dr. C. R. Rao and Dr. Martin Davidson, who read the book and made several valuable suggestions for improvement.

Thanks are also due to Mr. M. J. Moroney, whose frank and forthright criticism of some earlier drafts of the book helped to remove a number of faults, and to Sardar Khushwant Singh, who read the book to spotlight the difficulties of the enquiring layman and thus acted as a vigilant censor on his behalf.

FOREWORD

THIS is neither a text-book nor a general exposition of mathematics. It is an explanation of certain extremely useful branches of mathematics, some of which are little known to non-specialists. Though originally only of academic interest, they have been developed to tackle problems which proved unsolvable by the old mathematics. Some general knowledge of them is essential to the understanding of modern science, almost every department of which has its own special mathematical technique. This mathematisation of science is quite a recent phenomenon, for, until about the close of the nineteenth century, most branches of science were largely descriptive. It is true that the physical sciences, such as physics and astronomy, did use a good deal of mathematics, but even in these sciences one could get along and often make useful contributions without it. In fact, some of the most significant contributions to these sciences were made by non-mathematicians. Nowadays, even descriptive sciences, *e.g.* biology, zoology, genetics, psychology, neurology, medicine, economics, philology, *etc.*, have begun to employ elaborate mathematical techniques. The mathematics used is not always difficult, but it is often unfamiliar even to people who have had some mathematical training at a university. This is inevitable because the mathematics used is of recent origin and has not yet found its way into the school and college curricula.

One consequence of this development is that science is becoming the exclusive preserve of specialists. This is particularly unfortunate as science is the agency whereby our society has been changed in the past and will be changed still more rapidly and profoundly—perhaps within our own lifetime. It is dangerous to let this knowledge remain in the hands of a small group of specialists, however gifted, for the potentialities of science for good and evil are infinite. If this danger is to be averted, everyone, including you and I, must make a serious effort to understand what contemporary scientists are doing; but this understanding is impossible without some insight into the ideas of modern mathematics.

Fortunately, the need for popularising mathematical ideas is now universally recognised. More than twenty years ago an excellent popularisation of mathematics appeared in Professor Lancelot Hogben's *Mathematics for the Million*. This book deals with mathematics as it developed till about the middle of the eighteenth century. There is need for a similar popularisation of mathematical ideas that have come into being during the last two hundred years and are now proving so fruitful in genetics, economics, psychology, evolution, *etc.*, as well as in physics, astronomy and chemistry. The task will probably require many hands, if only for the reason that no single person can hope to master more than a few branches of the subject.

In this book I have tried to give a popular and, I hope, a not-too-inexact exposition of some major mathematical ideas that have been invented during the past two centuries or so. In explaining these topics, I have assumed the reader to have some knowledge of elementary mathematics such as could be obtained from any text-book on school algebra and geometry. I have tried to show how the most fruitful of these newer ideas arose as a result of man's impulse to mould his environment according to his heart's desire, and that where this impulse has been lacking mathematical progress has been stunted.

Although some of the theories explained in this book are still controversial, I have often taken 'sides' in presenting the different points of view, preferring to present things as I see them rather than as they might appear to an imaginary observer. However, the reader should find the treatment on the whole sufficiently unbiased for him to judge the issues for himself.

Calcutta JAGJIT SINGH

1

THE NATURE OF MATHEMATICS

WE learn about the universe around us by experience and observation on the one hand, and by thought and deductive reasoning on the other. Although in practice we get most of our knowledge by continually combining observation with deduction, it is possible subsequently to formulate certain types of knowledge by pure deduction starting from a set of 'axioms'—that is, statements accepted as true without proof because we feel that their truth is self-evident. The classic example of deductive method is school geometry, where we postulate certain definitions and axioms concerning points, lines, *etc.*, and deduce a logical chain of theorems concerning lines, angles, triangles, and so forth. The great advantage of the deductive method is the certainty of its conclusions. If there is no fallacy in our reasoning the conclusions must be correct—unless there is something wrong with our axioms. But the question whether the axioms chosen at the outset are valid is a difficult one. Certain axioms, which appear obvious to some, may seem clearly false or at least very doubtful to others. A familiar example of this kind is Euclid's axiom that parallel lines never meet. Mathematicians now recognise that it cannot be accepted as self-evident even though schoolboys are still taught geometry as if it were true. Indeed, the examination of the validity of any given axiom system is such a vexed question that some people propose to cut the Gordian knot by claiming that 'pure' mathematics is merely concerned with working out the consequences of stated axioms with no reference whatever to whether there is anything in the real world that satisfies these axioms.

Further support to this view was lent by the profound and penetrating studies of the foundations of Euclidean geometry towards the close of the nineteenth century. They revealed that geometrical proofs depend mainly on diagrams embodying properties which we accept as part of our equipment without including them in the axioms. Consequently an attempt was made to prove geometrical theorems without using the meanings of geometrical terms—like points and lines—as understood in everyday speech, but only their *relations and properties* as explicitly stated in the initial axioms. It was declared that 'if geometry is to be deductive, the deduction must everywhere be independent of the *meaning* of geometrical concepts, just as it must be independent of diagrams; only the relations specified in the axioms employed may legitimately be taken into account.'

But the insistence that the proof be independent of the meanings of the terms used, and employ only their mutual relations as explicitly stated in the basic axioms, did not mean that these terms were to be 'meaningless'* —only that they should have no specific reference to any particular thing. The terms of the axiom system should remain deliberately undefined— that is, free from association with any specific thing so that they become pure symbols formally related to one another in certain ways embodied in the axioms of the system.

There are two advantages in adopting such a procedure. First, by cutting out all associations that otherwise cling to terms used in the ordinary meanings of everyday speech, we eliminate the tendency to use meanings and relations other than those expressly stipulated in our axiom system— a tendency so prominent in Euclidean proofs. Second, the axiom system acquires a generality otherwise impossible. For example, it becomes possible to encompass within the single framework of an axiom system concepts, such as *group*† and *abstract space*, appearing in seemingly unrelated branches of mathematics. These very valuable gains are not to be despised. But the systematic draining of all meaning and content from the terms of our discourse, and thus turning them into pure symbols, means that mathematical proof becomes a sort of game with symbols. In a somewhat oversimplified form this is the view advocated by some mathematicians. Instead of playing-cards, dice, or pawns, bishops, rooks and knights, we may start with a collection of symbols such as $\sim$, $-$, $\times$, $+$, $=$, *etc.*, and a set of 'rules' or 'axioms' according to which they may be combined. We then proceed to play a game which consists of arranging these symbols to form 'meaningless' expressions according to the given 'rules'. We could change the symbols or the rules or both in any arbitrary manner we liked; the result would always be 'pure' mathematics.

Like all games this game of manipulating paper marks, that is, 'pure' mathematics, has, according to its present-day exponents, no ostensible object in view except the fun of playing it and playing it well. If men, nevertheless, find its results of great practical utility in their daily lives, that is not its *raison d'être*. Its sole function has been and should be merely to divert the human mind by the 'elegance' and 'beauty' of its expressions, irrespective of their utility. It is, no doubt, possible to argue in favour of this view. In fact, G. H. Hardy has done so in his charming little book, *A Mathematician's Apology*. His argument seems to rest on a distinction between what he calls 'school' or 'Hogben' mathematics, that is both

* This usage of the word must be distinguished from that of everyday speech as a term of disparagement. The word 'meaning-free' comes closer to the sense in which it is used here.

† See Chapter 7.

'trivial' and 'dull' but has considerable practical utility, and 'serious' mathematics that alone is the 'real' mathematician's delight despite its remoteness from everyday life.

It is evident that even those who hold this view cannot boast too seriously about the uselessness of their work and the 'meaninglessness' of the symbols and rules of their game. They do not go so far as to express dismay when their work turns out to be useful! Moreover, they are sufficiently sane to allow that a 'certain sense of fitness of things' should at any rate forbid a 'wild overturning of the law and order established in the development of mathematics'.

Mathematics is too intimately associated with science to be explained away as a mere game. Science is serious work serving social ends. To isolate mathematics from the social conditions which bring mathematicians (even of the Game Theory school) into existence is to do violence to history. Hogben and other writers have shown how great mathematical discoveries and inventions have throughout history been rooted in the social and economic needs of the times. Most books take us little beyond the eighteenth century in tracing this connection. A mistaken view has grown up in certain quarters that modern mathematics, particularly during the last 150 years, is a 'free creation' of the human mind, having little or nothing to do with the technological and social demands of the time. If science and technology have been able to make use of such mathematical inventions as, for example, the tensor calculus in Relativity and the matrix theory in Quantum physics, that in no way influenced their creation. Nevertheless, as will appear in the sequel, there is a close tie-up between the practice and theory (which has largely remained mathematical even up to the present time) of science and technology. Thus, while the empirical practices of the eighteenth-century mechanical engineers from Savery to Watt led to Thermodynamics, the 'pure' theory of Faraday and Maxwell paved the way for the practical inventions of Edison and Marconi. As Leonardo da Vinci remarked, science is the captain and practice the soldiers: both must march together. One reason why Einstein's Relativity theory has not yet advanced very much beyond the stage it reached during the second decade of the twentieth century may be the fact that it has made no practical difference in the calculation of the astronomical tables in the *Nautical Almanac*. If, at some future date, we are able to undertake interplanetary voyages, relativity might find a field of applications for want of which it has languished.

In spite of the 'game' theory, mathematics is still largely inspired by contemporary social, technological and scientific demands, as in the new 'pure' mathematics that has been created to deal with the problems of cosmic rays, stellar dynamics, stochastic processes and cybernetics. The present-day needs of science and technology for speedier methods of calculation by means of electronic machines may well inaugurate as great a mathematical revolution as the Hindu invention of zero and the positional system of writing numbers.

Far from despising utility and practical applications, the early pioneers of modern science and mathematics—men like Huygens, Newton and Leibnitz, to name only a few—cultivated them mainly for the advancement of technics. Is it any wonder then that the founders, patrons and some of the first scientists of such scientific academies as the Italian *Academie del Simento*, the English Royal Society or the French *Academie des Sciences*, were kings, nobles, courtiers, magnates and city merchants? The idea of a machine as the demiurge of a new heaven on earth was so uppermost in their minds that the very first standing committee of the Royal Society had for its terms of reference the 'consideration and improvement of all mechanical inventions'.

What, then, has given rise to the recent idea that mathematics is a game, a *jeu d'esprit* or 'free creation' of the mind divorced from the practical problems of daily life? It is the fact that the intimate connection between mathematics and reality is lost sight of in the abstract logical schemes which a mathematician constructs, though these always embody certain essential features abstracted from some sphere of reality. These logical schemes created by mathematicians do often *look* like games of manipulating symbols according to certain rules. This, however, does not mean that arithmetic, geometry, algebra, the calculus, *etc.*, arose by someone constructing these theories as games played according to some rules. Quite the contrary. They arose as abstractions from concrete applications, though their *subsequent logical* formulation may appear like games played with symbols. The authors of the game theory are, of course, aware of this distinction between the historical genesis of mathematical knowledge and its subsequent logical formulation. But when they claim that mathematics is a game they seem to confuse the means of expressing mathematical truths and the mathematics itself. (We shall deal with this subtle question more fully in the last chapter.)

2

NUMBER AND NUMBERS

When Ulysses had left the land of the Cyclops, after blinding Polyphemus, the poor old giant used to sit every morning near the entrance to his cave with a heap of pebbles and pick up one for every ewe that he let pass. In the evening when the ewes returned, he would drop one pebble for every ewe that he admitted to the cave. In this way, by exhausting the stock of pebbles that he had picked up in the morning he ensured that all his flock had returned.

The story is apocryphal, but this is precisely what the primitive shepherd did with his sheep before he learnt to count them. This also is not very far from what a modern mathematician does when he wants to compare two infinite collections, which *cannot* be counted in the ordinary way. However, the important difference between the two is that while the former used this tallying process without knowing what he was doing, like M. Jourdain speaking prose, the latter uses it with knowledge and insight. He thus acquires certain powers, otherwise unattainable, such as the power to count the uncountable. We shall see later (in Chapter 5) how the mathematician, by refining the primitive shepherd's practice, has succeeded in accomplishing this and other seemingly paradoxical feats. Meanwhile, we may note the theory behind the shepherd's practice. This theory is based on the fact that if the individuals of a flock can be matched, one by one, with those of a heap of pebbles so that both are exhausted together, then the two groups are equal. If they are not, the one that gets exhausted earlier is the lesser.

What gives this matching process its great power is that it can be applied universally to all kinds of aggregates—from collections of ewes and pebbles to those of belles and braces, apples and angels, or virtues and vipers. Any two aggregates whatever can be matched so long as the mind is able to distinguish their constituent members from one another.

Gradually men formed the notion of having a series of standard collections for matching the members of any given group or aggregate. One such series consisted of the ten different collections formed by including one or more fingers of their two hands. All collections, which could, for example, be matched on the fingers of one hand were 'similar' in at least one respect, however they might otherwise differ among themselves. They were, as we now say, all equal. These standard collections were then given names—

One, Two, Three . . . *etc.* This is the social origin of the practice of counting. Thus, when we now say that the number of petals in a rose is five, all that we mean is that if we start matching the petals one by one with the fingers of one hand, the members of both the collections are exhausted simultaneously. By long practice in handling the abstract symbols 1, 2, 3 . . . we are liable to forget that they are only a shorthand way of describing the result of an operation, *viz.*, that of matching the items of an aggregate with those of some set of standard collections that are presumed to be known. The process is so habitual that it usually escapes notice. This has caused endless confusion in the past, when, for long centuries, even learned men failed to understand the nature of number, particularly when they began to handle negative and imaginary numbers. If we keep in view the fact that whole numbers or integers are a mere shorthand for describing the result of a matching process, in which one of the collections is presumed to be known, we shall avoid a lot of trouble in understanding the nature of more sophisticated types of numbers in mathematical literature.

We have seen that originally man formed his standard collections for counting with the fingers of his hands. In the beginning this sufficed, there being no occasion to budget for atomic piles, armament races, refugee reliefs, or Marshall Aid. But presently, even in the days of the river-valley civilisations of antiquity, the needs of armies, taxation and trade gave rise to collections which could hardly be matched on the fingers of the two hands. What could man do about it? He could use the marks on his fingers instead of the fingers themselves for the purpose; but even so he would not have enough of them. But as the matching process was independent of the nature of the members constituting the collections, it did not matter whether he formed them by means of fingers or finger-marks or anything else. So he conceived the idea of generating a new standard collection from one already known by mentally adding just one more item to it. And as the process could be repeated indefinitely, he produced an unending succession of standard collections some one of which sufficed to match any given collection, however large. Thus to the original idea of matching or tallying was grafted another—that of order—in virtue of which relative rank is given to each object in the collection. Out of the union of the two arose the idea of integral number—an unending, ordered sequence of integers.

In matching two collections we have hitherto considered them as mere crowds of individuals without any internal order between themselves. The concept of number as the characteristic of a class of similar collections evolved from the practice of matching unordered aggregates (*e.g.* using any pebble in Cyclops' hand for tallying any ewe in the fold) is known as the *cardinal* number. However, we can also conceive of ordered aggregates, such as soldiers in a battle-array or ewes arranged in a straight file, in which every element has a rank or place. When we match such ordered aggregates and conceive of number as the characteristic of a class of similar ordered aggregates, which exhaust themselves together in a matching process, it is known as the *ordinal* number. Primitive man used both these concepts without making any distinction between them. When he used pebbles like Cyclops, he was using *cardinal* numbers; but when he used his fingers he probably used them in a definite order, possibly first his right-hand thumb, then the index, middle, ring and little fingers. In the latter case, he was using *ordinal* numbers. The distinction between the two types of numbers is somewhat subtle and was not noticed by mathematicians themselves till about the end of the nineteenth century. Fortunately, it is of no great importance for all ordinary purposes and may be ignored. The sole justification for introducing what may appear to some a pedantic distinction is the importance it assumes in the theory of transfinite numbers.*

Now, quite early in life we are taught the technique of adding and multiplying the integers. Underlying this technique are certain general laws of addition and multiplication. Though known by high-sounding names they merely express in symbolic language just one commonplace fact of everyday experience. That is, that it makes no difference in what order you add the various sets of objects. Thus whether you buy two books on the first occasion and three on the second or vice versa, your total purchase remains the same. We express facts like this by the formula $2 + 3 = 3 + 2$. This formula is a particular case of a more general law, the *commutative law of addition*, which requires that the sum of any two integers such as a and b is the same in whatever order we may choose to add them. In symbols,

$$a + b = b + a \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Next, there is the *associative addition law*, expressed by the equality

$$(a + b) + c = a + (b + c)^{\dagger} \quad . \quad . \quad . \quad . \quad (2)$$

Since multiplication is only reiterated addition, there are naturally also corresponding *commutative* and *associative multiplication laws*, the counterparts of the addition laws. Thus:

$$a.b = b.a \quad . \quad (3), \textit{ corresponding to } (1)$$

and

$$(ab)c = a(bc) \quad . \quad (4), \textit{ corresponding to } (2)$$

Finally, we have the *distributive law*

$$a(b + c) = ab + ac \quad . \quad . \quad . \quad . \quad . \quad (5)$$

* See Chapter 5.

† The brackets here simply mean that the numbers within them are to be added first.

In the following sections we shall see how the number system of integers can be extended to include other types of numbers like negative, fractional and irrational numbers. It will also be shown that the above five fundamental laws hold equally well for these more complicated types of numbers. Further, they remain valid even for sets of symbols which are not numbers but behave *like* numbers in some respects. We shall cite several instances of such sets later.* But the fact that such sets of symbols are possible has enabled modern mathematicians to make truly amazing feats of abstraction. We can, for instance, picture a set of 'elements' a, b, c, d . . . about which we assume nothing except that they obey the five fundamental laws of arithmetic. Starting from this assumption we can prove a number of theorems about them. These theorems will hold, not only for numbers—whether integers, fractions or irrationals—but also for a much wider class of symbols which includes these numbers as a special case. In other words, we are able to subsume the properties of a vast variety of elements under one generic form that applies to them all. This is the method of abstraction which is the very life-breath of modern mathematics.

Indeed, the main difference between ancient and modern mathematics is just this. In its relentless drive towards greater abstraction it refuses to tie up the symbols of its discourse to anything concrete, so it cannot make much use of the meanings of these symbols. How could it, indeed—since it refuses to give them any meanings, or at any rate keeps them as meaning-free as possible? Mathematics thus has to base itself more on the mutual relations of abstract symbols, as embodied in laws like the five fundamental laws of arithmetic, than on the meanings of these symbols. That is why some people call ancient mathematics '*thing-mathematics*', meaning concrete, and modern mathematics '*relation-mathematics*', meaning abstract.

This difference does not imply that modern mathematicians are cleverer or more imaginative than their forefathers. It is largely the outcome of changes that have since occurred in the mode of civilized living. In the ancient world, where material production was largely for use and barter, things had not yet become completely metamorphosed into abstract embodiments of sale or money value. This could occur only with the change from a productive system for use or barter to a money economy producing commodities. In such a social system all qualitative differences between commodities are effaced in money and one begins to think in terms of their money values as—four talents, four livres or four guineas. From this it is but a step to think of the number 'four' as dropping its material crutches and coming into its own as an abstract conceptual symbol for the common quadruplicity of them all. But this is by no means

* See pages 21 and 146.

the end. A modern mathematician is able to carry on the process of abstraction much farther. All that he needs is a system of elements obeying some scheme of abstract laws like the five laws of arithmetic cited above.

You may wonder whether these abstract schemes about phantom entities obeying phantom laws are of any use at all. They are indeed. As science advances, sooner or later a stage is reached when it has to reckon with what is sometimes called interphenomena, that is, phenomena beyond the limit of direct observation. Thus no one can see what actually happens inside a star, an atom, a gene, a virus, an amacrine cell or an ultramicroscopic speck of nerve fibre. And yet a scientist must somehow figure it out if he is to give an intelligible account of perceptible phenomena. One way of doing it is to adopt some abstract scheme of a mathematician's fancy and see where it leads us. It may happen that it enables us to predict some observable phenomena capable of direct observation. If we do succeed in observing the predicted phenomena, we may be sure that the abstract scheme does embody at least some features of the interphenomena under study. Surprising as it may seem, the method actually works. It is by the use of such abstract mathematical schemes that scientists have been able to fathom what happens in the interiors of stars and atoms.

Stars and atoms may seem very remote, but they have now begun to influence our everyday lives very directly, for their study has revealed new sources of power which, in the case of atoms, we may use for war or peace as we may choose. For instance, thanks to these studies we can now imitate, though in a rudimentary manner, the cosmic processes at work in the solar interior and construct here on earth those miniature suns, the H-bombs, which threaten to wipe the human race out of existence. And yet the same theory which has produced the H-bomb is also potentially capable of putting, as it were, sunshine on tap for the advancement of technics, civilisation and life.

* * * *

As we have seen, early man's need to compare discrete groups, such as flocks of sheep, herds of cattle, fleets of vessels and quivers of arrows, gave rise to the concept of the integer. His other needs had even more fruitful consequences. For instance, he wanted to know whether the milk-yield of his cattle was rising or falling, or to make sure that no one encroached on his field at harvest time. Here he was faced with *continuously* varying quantities that could not be counted like the discrete objects of a group, such as eggs in a basket. Nevertheless, he found a way of reducing the problem of quantizing these continuous magnitudes to that of counting a discrete group. Thus, he took a standard yard-stick and counted the number of such sticks, which, placed end to end, covered the entire length of the field from one extremity to the other. The continuous length thereby became a discrete group of equal yard-sticks. Or he took a standard vessel and poured out of his milk-yield as many vessel-fuls as he possibly could. The continuous quantity—*i.e.*, milk-yield—was thus changed into a discrete collection of standard-sized vessels. In this way continuous quantity became discrete and could be counted. But the solution had its awkwardness. For, in reducing the length of a field to a discrete group of yard-sticks or unit-measures, it might happen that a residue was left. For instance, the yard-stick might cover a given length 200 times and then leave a residue smaller than the yard-stick. What was he to do with it?

There were two ways of handling it, namely to find a unit that leaves no residue or to ignore it altogether. The first method was used in measuring time when men divided the duration of daylight into twelve equal hours. The hour was thus a variable unit, for a duration of one hour at the time of summer solstice, for instance, was not the same as that at the time of vernal equinox. At a time when there were no Hours of Employment Regulations, overtime wages, payment by the hour, *etc.*, this meant no social inconvenience. It was otherwise with measuring the lengths of fields. For it would have been necessary to discover by trial and error a unit that would cover each length exactly without leaving a residue. Even if he found one it would most probably be of no use for measuring another length, for which it would be necessary to seek another unit. If, therefore, residues were to be eliminated by this method, almost every length would have needed a special unit of its own—very much like Chinese writing, which has a separate ideograph for every word in its vocabulary. With such a medley of different units, not only the calculus of lengths would have been more complicated than Chinese writing but few lengths could have been compared with each other. A fixed standard of length, as also of weight and volume, was therefore a prime necessity of social intercourse. Consequently, while all the early civilisations adopted variable units of time, the units of lengths, weights and measures were fixed. As early as 2300 B.C. the Sumerian Hammurabi, for example, issued edicts fixing these standards.

But if these units were to remain invariable, the problem of the residue had to be solved. Now it would be a mistake to imagine that early man became 'residue-conscious' overnight. Far from it. Actually this residue-consciousness took whole millennia acoming. Too often and too long men tried to ignore the residues and were content to make do with approximations, even ludicrous approximations, such as the value of $\pi = 3$ adopted in the *Book of Kings*, in *Chronicles** and by the Babylonians. Neverthe-

* 1 Kings vii, 23; 2 Chronicles iv, 2.

less, the problems of trade and the administration of the vast revenues of temples, city states and empires, kept in the forefront the mathematical problem of the residues. Even then our present insight into the nature of fractional numbers, which were created to solve it, was gained with the greatest difficulty. The Egyptians, who treated only fractions with the numerator 1, never understood that fractions were amenable to the same rules as integers. The Babylonians, who had begun to deal with fractions as early as 5000 B.C., did not acquire complete mastery of fractional numbers until 2000 B.C. We can well appreciate their difficulty: they faced problems unprecedented in human history. To solve them they had to find their way about in an utterly strange and uncharted domain. No wonder that, like the early navigators, they often reached their goal by the longest and most devious route. And yet it is easy, retrospectively, to state the residue problem and its solution. The problem of the residue is simply this:

If the fixed unit yardstick does not happen to go into the length to be measured an exact number of times, how are we to measure the residue? The solution is equally simple. Divide the unit yardstick itself into a number of aliquot parts and then measure the residue with it. For instance, the yard may be divided into 3 sub-units, each one foot long, and we may try to measure the residue with the sub-unit, the foot. It may be that the new sub-unit will lie along the residue from end to end exactly twice. We thus have a sub-unit which covers the residue twice and our original unit—the yardstick itself—thrice. This process gives us a number couple, *viz.* 2 and 3, which can be used as a measure of the residue. We may write it as $\frac{2}{3}$ as our Hindu ancestors did, or $\frac{2}{3}$ or 2/3, as in modern text books, or as (2,3) as some learned people might advocate; it matters little. The important point is that it is a shorthand for an operation just as the single integer by itself is a symbolical way of describing the result of a matching process. A number pair like $\frac{2}{3}$ or (2,3) says what in ordinary language would have to be expressed somewhat as follows:

'If you take a sub-unit that divides the fixed unit, say a yard, exactly thrice, and use it to span the residue in question, it will go into the residue exactly twice'.

The problem of measuring residues is, therefore, merely the problem of finding *some* sub-unit that will cover exactly both the residue and the fixed unit. Although, as we shall see later, strictly speaking it is not soluble in all cases, a solution that is good enough for all practical purposes can always be found. For, if we take a sufficiently small sub-unit of the original unit, it will either cover the residue exactly or, at most, leave a remainder less than the chosen sub-unit. But as the sub-unit can be made as small as we please, the remainder, if any, will always be smaller still.

Having found a way to measure residues or fractions, as they are generally called, it was necessary to devise methods of adding and multiplying them. In time, general rules for adding and multiplying fractions were framed. They are the well-known rules we all learnt at school, *viz.*:

Addition rule: $(a, b) + (c, d) = (ad + bc, bd)$,
Multiplication rule: $(a, b)(c, d) = (ac, bd)$,*

You could, if you were in doubt, even 'prove' these rules by strict logic. All that you need do is to recall the meaning of number pairs (a, b) and (c, d) and the way to add or multiply them would be clear. But the point that is of greater interest is that these two rules for adding and multiplying fractions show that fractions too obey the same five fundamental laws of arithmetic as the integers. This too can be proved logically.

*   *   *   *

If you operate with the first three integers 1, 2, 3, you may easily verify that you can combine them in pairs in $3 \times 3 = 9$ ways. These nine ways lead to the nine fractional numbers:

$$(1, 1), (1, 2), (1, 3)$$
$$(2, 1), (2, 2), (2, 3)$$
$$(3, 1), (3, 2), (3, 3)$$

Likewise, four integers produce $4 \times 4 = 16$ fractions and five integers $5 \times 5 = 25$ fractions. From this you may easily infer that with N integers you can manufacture $N \times N = N^2$ fractions. This would seem to show that the set of fractional numbers is vastly more numerous than that of positive integers. But the set of positive integers is a never-ending or infinite set. If we denote this infinity by the usual symbol ∞, the infinity of fractional numbers would appear to be the much bigger infinity $\infty \times \infty = (\infty)^2$. And yet if, in the manner of the Cyclops, you started matching the infinite set of fractional numbers with the infinite set of integers, both the sets would be exhausted together—provided one could speak of exhausting inexhaustible or infinite sets! In other words, the two infinite sets of integers and rational fractions are exactly 'equal'. This is, no doubt, paradoxical. We shall explain this paradox in Chapter 5.

*   *   *   *

If early man noticed that one herd of cattle could be 'added' to another and he thus formed the notion of 'addition', he also performed the reverse

* (a, b), (c, d) are here used to denote the fractions a/b, c/d.

operation—that of taking some cattle out of his herd as, for instance, for the purpose of bartering them for other goods. This is the origin of 'subtraction', the inverse of addition. Similarly, multiplication, which is only a reiterated addition, gave rise to its inverse, 'division', a reiterated subtraction. At first, these new operations caused him some confusion. For, while he could always add and multiply *any* two integers, he could not always perform the inverse operations. Thus, he could add *any* two herds of cattle, but he could not take out, say, fifty cows from a herd of only forty. Division, too, must have worried him at times, and he must have often wondered whether a division of, say, seven by two is possible at all. Like most children beginning to learn arithmetic, he, too, must have felt that there isn't a '*real*' half of seven.

Nevertheless, for two reasons early man had less trouble with division than with subtraction. First, he could always divide one integer, say 7, by another, say 2, and supplement the result by adding that the division is not 'exact' and leaves a *remainder*. Second, even if he had to divide a smaller number, say 5, by a larger, say 7, he could interpret the result as a number pair (5, 7)—the fractional number that he had already devised for measuring continuous magnitudes. But if he was asked to subtract, say 7 from 5, he was quite befogged. To make this magic possible, he had to wait for the rise of a banking system with an international credit structure, such as came into being in the towns of Northern Italy (particularly Florence and Venice) during the fourteenth century. The seemingly absurd subtraction of 7 from 5 now became possible when the new bankers began to allow their clients to draw seven gold ducats while their deposit stood at five. All that was necessary for the purpose was to write the difference, 2, on another side of the ledger—the debit side.

Although the attempt to resolve the difficulties of awkward divisions and subtractions did lead to the recognition of fractional and negative numbers, the realisation that they arose from the limitations of the integral number system itself, and could only be overcome by suitably extending that system, came much later. Thus, suppose we are given only the unending sequence of positive integers 1, 2, 3, ... We can clearly add any two of these integers, their sum being itself a positive integer. But we cannot always perform the inverse operation of subtraction. For instance, while we can subtract five from seven we cannot subtract seven from five. If we want to ensure that subtraction of one integer from another be as freely possible as addition, we must extend the number system of positive integers to include negative integers, so as to form a doubly unending set of positive and negative integers:

$$\ldots -4,\ -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,\ 4, \ldots$$

Only when we operate with such an extended set can we perform the operation of addition and its inverse subtraction on *any* two numbers without any restriction whatsoever. In other words, to make subtraction universally possible the system of positive integers must be extended to include negative integers too. This has as a consequence that to every positive integer, such as a, there corresponds its negative or inverse, $-a$, which belongs to the same set and is such that the sum $a + (-a)$ is zero. This fact may also be expressed by the statement that the equation $a + x = 0$ has always a solution $x = -a$ (also belonging to the set). If we consider the particular case of this equation when $a = 0$, we find that x too is zero. It therefore follows that the number zero of the set is its own inverse. It is called the '*identity element*' *of the set for addition*. We call it the *identity* element as its addition to any integer leaves the latter unaltered. If we imagine that the relation between an integer and its inverse is like that of an object and its mirror image, then the identity element zero is like the reflecting surface, which is its own image.

Our first extension of the number system is thus the doubly unending set of positive and negative integers complete with the identity element zero:

$$\ldots -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots$$

Such a system of positive and negative integers, including zero, is known as an *integral domain*. It is not possible to extend the integral domain any farther by means of addition and subtraction alone. No matter which two numbers of the domain we may add or subtract, we shall always end up with a positive or negative integer belonging to the integral domain. We may say that the integral domain is *closed* under addition and subtraction because the way to further extension of the domain by performing these two operations is blocked.

Just as we had to extend the system of positive integers to include negative integers in order to make subtraction universally possible, a similar consideration with regard to division leads to a further extension of integers to include rational fractions. For, so long as we work with integers, we can always multiply any two of them, but the inverse operation—division—is not always possible. To make division between any two integers as universally possible as multiplication, we have to extend the number system to include rational fractions, that is, number pairs devised to measure residues. In other words, the set of all integers must be extended to include all rational fractions so that each number a (other than zero) of the set has an inverse $1/a$ belonging to the set with respect to multiplication. We may express the same thing by saying that the equation $ax = 1$ has a solution $x = 1/a$ belonging to the set for all a's not equal to zero. If we consider the particular case of this equation when $a = 1$, we find that x is also 1. It therefore follows that the number 1 of the set is its own inverse. It is called the "identity element"* for multiplication, since multiplying it by any number leaves the latter unaltered. Such a system which includes all positive and negative integers as well as fractions is called a *field.*

With the construction of the field of rational numbers our second extension of the number system is in a way complete. It permits us to perform on *any* two numbers of the system not only addition and its inverse, subtraction, but also multiplication and its inverse, division, and express the result by a number belonging to the rational field. No further extension of the field is possible by performing any of these four arithmetical operations on any two numbers of the rational field. No matter which two numbers of the rational field we may add, subtract, multiply or divide, we shall always end up with a number belonging to the rational field. In other words, the field of rational numbers is *closed* under all the four arithmetical operations. But even so the number system still remains incomplete in some ways. It can be shown that certain magnitudes like the diagonal of a unit square cannot be measured by rational fractions. To measure the diagonal we have to find a sub-unit that goes an exact number of times into the side as well as the diagonal. Suppose, if possible, there is such a sub-unit which divides the side m times and the diagonal n times. Then the length of the diagonal is given by the number pair or fractional number n/m. We may assume that m and n are *not* both even, for if they were, we could cancel out the common factor 2 from the numerator and denominator till one of them became odd. Now, the lengths of the two sides of the square AB, BC, are both equal to unity, and that of the diagonal is n/m (see Fig. 1). But in a right-angled triangle like ABC,

$$AB^2 + BC^2 = AC^2.$$

In other words,

$$2 = n^2/m^2$$

or

$$n^2 = 2m^2.$$

But it can be proved that no two integers can satisfy this equation unless they are both even—a possibility which has been expressly excluded by our hypothesis. It is therefore impossible to find a rational number to measure the diagonal length AC.

The discovery of magnitudes which, like the diagonal of a unit square, cannot be measured by any whole number or rational fraction, that is, by means of integers, singly or in couples, was first made by Pythagoras some 2500 years ago. This discovery was a great shock to him. For he was a

* Note carefully that the number 1 plays the same role with regard to multiplication as the number zero with regard to addition. (See page 14.)

number mystic who looked upon integers in much the same reverential spirit as some present-day physicists choose to regard Dirac's *p*, *q* numbers, *viz.* as the essence and principle of all things in the universe. When, therefore, he found that the integers did not suffice to measure even the length of the diagonal of a unit square, he must have felt like a Titan cheated

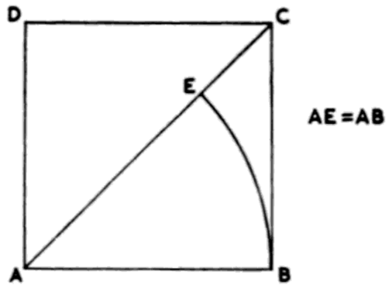

FIG. 1—The diagonal and the side of the square are incommensurable.

by the gods. He swore his followers to a secret vow never to divulge the awful discovery to the world at large and turned the Greek mind once for all away from the idea of applying numbers to measure geometrical lengths. He thus created an impassable chasm between algebra and geometry that was not bridged till the time of Descartes nearly 2000 years later.

Nevertheless, there is no great difficulty in measuring the length of the diagonal by an extension of the process that gave rise to fractions. Suppose we use the unit stick to measure the diagonal AC of the square. It goes once up to E and leaves a residue EC (see Fig. 1). Its length is therefore $1 + EC$. Now suppose we use a sub-unit which goes into the original unit ten times, in order to measure EC. We shall find that it will cover EC four times, leaving again a small residue. This gives us a closer measure of the diagonal length, *viz.*, $1 + 4/10$ *plus* a second residue. To obtain a still closer estimate, we try to measure the second residue with a still smaller sub-unit which divides our original unit into, say, 100 equal parts. We shall find that it goes into the second residue just once but again leaves a third residue. Our new estimate is thus $1 + 4/10 + 1/100$ *plus* a small residue. To measure this last residue, we can again use a still smaller sub-unit, say, one-thousandth part of the unit, and observe how many times it covers it. We shall find that it covers it four times and still leaves a residue. The diagonal length then is $1 + 4/10 + 1/100 + 4/1000$ *plus* a residue.

Now, the important point is that no matter how far we go, the residue always remains. For suppose, if possible, we were left with no residue when we began to use a scale equal to, say, a thousandth part of our unit. This would mean that a scale that covered the unit 1000 times would cover the diagonal 1414 times exactly. In other words, it would be measured by the fraction 1414/1000, but this, as we saw (page 15), is impossible. Thus, while we may be able to reduce the residue to as small a length as we please by prolonging this process far enough, we can never hope to abolish it altogether. How shall we express, then, the length of our diagonal if we want to do it exactly? We can do so by the never-ending sum:

$$1 + 4/10 + 1/100 + 4/1000 + 2/10{,}000 + \ldots$$

Or, if this is too cumbersome, we may use the non-terminating decimal expression, 1·4142 . . ., which is only an abridged way of writing the same thing. A still shorter way of writing it would be 'square root of 2' or $\sqrt{2}$, which means that it is *some* number whose square is two. Numbers like $\sqrt{2}$ were called 'irrationals', as they did not appear to be amenable to reason; they escaped the number mesh cast by man to trap them. For instance, take $\sqrt{2}$ itself: if we do not wish to go beyond the first decimal place, the fraction 1·4 is a little too small, while 1·5 too large. By going to the second place, we can get a closer mesh, *viz.*, 1·41 and 1·42. The third decimal place gives us 1·414 and 1·415, fourth decimal place 1·4142 and 1·4143, and so on. If we square the end numbers of each of these meshes of the number-net we have

$$\begin{aligned}
(1{\cdot}4)^2 &= 1{\cdot}96 &< 2 &< (1{\cdot}5)^2 = 2{\cdot}25;\\
(1{\cdot}41)^2 &= 1{\cdot}9881 &< 2 &< (1{\cdot}42)^2 = 2{\cdot}0164;\\
(1{\cdot}414)^2 &= 1{\cdot}999396 &< 2 &< (1{\cdot}415)^2 = 2{\cdot}002225;\\
(1{\cdot}4142)^2 &= 1{\cdot}99996164 &< 2 &< (1{\cdot}4143)^2 = 2{\cdot}00024449;\\
&\ldots
\end{aligned}$$

We note two properties of the end numbers of this mesh system that we have created to trap the square root of 2. First, the left-hand end numbers, *viz.* 1·4, 1·41, 1·414, 1·4142 . . . . continually increase or at least never decrease. Second, no matter how far we go, the square of any number in the sequence always falls short of 2, though the difference continually decreases. Similarly the right-hand end numbers, *viz.* 1·5, 1·42, 1·415, 1·4143 . . . . continually decrease or at any rate never increase. Likewise, the square of any member always exceeds 2, though the difference continually diminishes the farther we go. We have here a process whereby we generate two sequences of fractional numbers which continually approach the square root of 2 from both below and above, although they never actually reach it. Whatever the degree of precision required in the estimate, we can always pick up two numbers, one from each sequence, which are sufficiently close together, and between which the desired square root lies. Can we, then, say that this hunt for the square root of 2 is not as 'perfect and absolute a blank' as that of Lewis Carroll's crew in the hunting of the Snark? In other words, can we assert that the square root of 2 'exists', in spite of the fact that it cannot be expressed as an ordinary fraction? The question is not entirely academic, for unless we admit its 'existence', our number system is not complete. Our number vocabulary is not rich enough to quantize certain magnitudes. Without the irrationals we should have no numbers for exactly measuring certain lengths, although we might have increasingly finer sets of approximations. Voltaire once remarked that if God did not exist, it would be necessary to invent Him. With still greater justification, the mathematician says that if the square root of 2 does not exist, it is necessary to invent it, and he invents it by writing $\sqrt{2}$. It is the unique number towards which the infinite ever-increasing sequence of fractional numbers 1·4, 1·41, 1·414, 1·4142 . . . continually tends without ever reaching it. In other words, it is the *ultima Thule* or *limit* of this sequence which we usually abbreviate as the non-terminating, non-recurring decimal, 1·4142 . . . .

Now it is no use inventing numbers unless we know how to combine them by addition, multiplication, *etc.* What do we mean by adding two irrational numbers like $\sqrt{2}$ and $\sqrt{3}$? We defined $\sqrt{2}$ as the *limit* of an infinite never-decreasing sequence of rational fractions, such as:

$$1{\cdot}4,\ 1{\cdot}41,\ 1{\cdot}414,\ 1{\cdot}4142,\ \ldots\ldots\ldots$$

Likewise $\sqrt{3}$ is the limit of another infinite never-decreasing sequence of fractions, namely:

$$1{\cdot}7,\ 1{\cdot}73,\ 1{\cdot}732,\ 1{\cdot}7321\ \ldots\ldots\ldots$$

The sum of $\sqrt{2}$ and $\sqrt{3}$ is merely the limit of the new infinite never-decreasing sequence formed by adding the corresponding terms of these two sequences *viz.*:

$$(1{\cdot}4 + 1{\cdot}7),\ (1{\cdot}41 + 1{\cdot}73),\ (1{\cdot}414 + 1{\cdot}732),\ (1{\cdot}4142 + 1{\cdot}7321),\ \ldots\ldots$$

Likewise, the product of $\sqrt{2}$ and $\sqrt{3}$ is the limit of the infinite sequence:

$$(1{\cdot}4)(1{\cdot}7),\ (1{\cdot}41)(1{\cdot}73),\ (1{\cdot}414)(1{\cdot}732),\ (1{\cdot}4142)(1{\cdot}7321),\ \ldots$$

Since fractional numbers are subject to the commutative, associative and distributive laws of arithmetic, so are the irrational numbers like $\sqrt{2}$ and $\sqrt{3}$, as they are defined as mere limits of infinite sequences of rational fractions. For instance, $\sqrt{2} + \sqrt{3}$ is the limit of the sequence

$$(1{\cdot}4 + 1{\cdot}7),\ (1{\cdot}41 + 1{\cdot}73),\ (1{\cdot}414 + 1{\cdot}732)\ \ldots . \qquad (1)$$

and $\sqrt{3} + \sqrt{2}$ is the limit of

$$(1{\cdot}7 + 1{\cdot}4),\ (1{\cdot}73 + 1{\cdot}41),\ (1{\cdot}732 + 1{\cdot}414)\ \ldots . \qquad (2)$$

Sequences (1) and (2) are obviously identical and consequently so also are their respective limits. In other words,

$$\sqrt{2} + \sqrt{3} = \sqrt{3} + \sqrt{2}.$$

The addition of irrational numbers to the field of rational numbers makes what is known as the *real number field.* It is the aggregate of all integral, fractional and irrational numbers, whether positive or negative. It is obvious that we can perform any of the four arithmetical operations on any two of its numbers and express the result as a number belonging to itself. This means that the real number field is also closed under the arithmetical operations. It might thus appear that our third extension of the number system is at last complete. But, as we shall see later, the real number field too is incomplete in some ways and needs further extension.

* * * *

As we have seen, starting with positive integers, the number domain was extended to cover the entire set of real numbers by the invention of negative numbers, fractions and irrationals. We shall see later how the idea of vectors and complex numbers grew out of real numbers, and that of quaternions and hypercomplex numbers out of vectors and complex numbers. With the invention of hypercomplex numbers the art of number-making seemed to have reached its acme, for any kind of number could be shown to be a particular case of some hypercomplex number. With the closing of the field of number-making, mathematicians returned to the integer from which they had started and opened another. In endeavouring to discover the essence of the integer they created a new subject—mathematical logic. By the first two decades of the twentieth century, they had succeeded in creating a mere mathematician's delight, and that to such a degree that it was in real danger of becoming what the Americans call 'gobbledygook'. Fortunately it was rescued from this disaster by the practice of electronic engineers, who applied it to produce new types of ultra-rapid automatic calculating machines employing all manner of electrical apparatus. With the invention of these new electronic devices it was possible to apply the abstract ideas of mathematical logic to advance the design of calculating machines far beyond the dreams of early pioneers like Pascal and Leibnitz or even Babbage.

The reason why mathematical logic has had such great influence on the art of numerical computation is that the calculus of reasoning is symbolically identical with the calculus of number. Since in logic we deal with statements or propositions which have some meaning, every such proposition is either true or false. Let us assign the truth value $T = 1$ when the proposition is true, and T = O when it is false. Every proposition such as A will then have a truth value T which may be either zero or one. If we have another proposition B, we can form a compound proposition from these two in two ways. First, we may produce a compound proposition S which is considered true provided *either* A *or* B is true. In this case S is the logical sum of A and B and the process of obtaining it is the analogue of numerical addition. Second, we may obtain another compound proposition P which is considered true if, and only if, both A *and* B are true. P is known as the logical product of A and B and the process of obtaining it is the counterpart of arithmetical multiplication.* For example, let A be the proposition 'Socrates drank the hemlock' and B the proposition 'Voltaire wrote *Gulliver's Travels*'. S, the logical sum of A and B, will then be the compound proposition:

S { *Either* 'Socrates drank the hemlock'
or 'Voltaire wrote *Gulliver's Travels*',

P, the logical product of A and B, will, on the other hand, be the compound proposition:

P { 'Socrates drank the hemlock'
and
'Voltaire wrote *Gulliver's Travels*'.

Since we know that in this case A is true and B false, then S will be true but P false. Consequently when the truth value of A is 1 and of B zero, that of S will be 1 and of P zero. In the same way we can easily work out the truth values of S and P, given those of A and B in any other case. In general, as mentioned earlier, for S to be true only *one* of the two constituents A and B need be true, whereas for P to be true *both* A and B have to be true. This rule suffices to evaluate the truth values of S and P as we shall now show.

Suppose both A and B are true so that the truth values of both are one. Since S is true when either A or B is true, the truth value of S is 1. This leads to the summation rule:

$$1 + 1 = 1.$$

If both A and B are false, then obviously their logical sum S too is equally false so that the summation rule now is:

$$0 + 0 = 0.$$

* *S* is also known as the disjunction of *A* and *B* and is written as $S = A v B$.
P is also known as the conjunction of *A* and *B* and is written $P = A.B$.

But if only one of the two, *viz.* A or B, is true, then S is also true, because S is true when either of them is true. This leads to the summation rules:

$$0 + 1 = 1, \; 1 + 0 = 1.$$

We may summarise these summation rules in the table of logical addition:

Logical Addition

+	0	1
0	0	1
1	1	1

To read the result of the addition of any two truth values, say 0 and 1, take the row 0 and the column 1; these are easily seen to intersect at 1. The same rule applies in reading all other tables described in this section.

Consider now the product proposition P. Since P is true only when both A and B are true, its truth value is 1 only when that of both A and B is 1. In every other case P is not true and therefore its truth value is zero. This leads to the product rules:

$$1 \times 1 = 1, \; 0 \times 1 = 0, \; 1 \times 0 = 0, \; 0 \times 0 = 0.$$

This may be summarised in the table of logical multiplication:

Logical Multiplication

×	0	1
0	0	0
1	0	1

We shall now show that these tables* of logical addition and multiplication are very similar to their counterparts of arithmetical addition and multiplication. Although we use the ten digits, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 to write numbers, this is merely due to the physiological accident that we have ten fingers. If we had had only eight, we might have worked with only eight digits *viz.* 0, 1, 2, 3, 4, 5, 6, 7. In that case what we now write as '8', '9' and '10' would be written '10', '11' and '12', respectively. A number like 123 written in this octonal notation would really be an abbreviation of $1(8)^2 + 2(8)^1 + 3(8)^0$ just as 123 in the decimal notation is a shorthand for $1(10)^2 + 2(10)^1 + 3(10)^0$. In the octonal notation, therefore, the number '123' would be $64 + 16 + 3 = 83$ in the usual decimal notation.

What notation we choose for writing numbers, whether decimal,

* With the help of these tables you may readily verify that the symbols 0, 1, though not numbers in the ordinary sense, yet obey the five fundamental laws of arithmetic. This is an instance of a set of 'elements' other than numbers satisfying the laws of arithmetic.

octonal or any other, is arbitrary. In principle we are free to make any choice we like. Of all the possible choices the simplest, though not the most familiar, is the binary notation in which we work with only two digits 0 and 1. It is remarkable that we can express any number whatever in the binary notation, using only these two digits. Thus, the number two would be written in the binary notation as 10, three as 11, four as 100, five as 101, six as 110 and so on. For 10 in the binary notation is $(2)^1 + 0(2)^0 = 2$ in the decimal notation. Likewise, 110 in the binary notation is $1(2)^2 + 1(2)^1 + 0(2)^0 = 4 + 2 + 0 = 6$ in the decimal notation. A binary 'millionaire' would be a very poor man indeed, for the figure 1,000,000 in the binary scale is a paltry:

$$1(2)^6 + 0(2)^5 + 0(2)^4 + 0(2)^3 + 0(2)^2 + 0(2)^1 + 0(2)^0 = (2)^6 = 64$$

in the decimal notation. Nevertheless, the binary notation is potentially as capable of expressing large numbers as the decimal or any other system. The only difference is that it is a bit lavish in the use of digits. Thus the very large number of grains of wheat which the poor King Shirman* was inveigled into promising his sly Grand Vizier as a reward for the latter's invention of chess could, in the binary notation, be expressed simply as a succession of sixty-four ones:

111, 111, 111, 111, 111, 111, . . . sixty-four times.

In the decimal notation of everyday use we should need twenty digits to write it.

The rules of ordinary addition and multiplication in the binary notation are

$$0 + 0 = 0; 0 + 1 = 1; 1 + 0 = 1; 1 + 1 = 10.$$

If we remember that while adding one to one, we get a 'one' which should be 'carried' to the next place, we can summarise the addition rules in the table of arithmetical addition:

Arithmetical Addition

+	0	1
0	0	1
1	1	1

* The allusion here is to the well-known legend of the Grand Vizier who asked for one grain of wheat in the first square of a chessboard, two in the second, four in the third, eight in the fourth, sixteen in the fifth and so on till the sixty-fourth square. The poor king never suspected till it was too late that the total number of grains required to fill the board in this manner would exceed the total world production of wheat during two millennia at its present rate of production!

Similarly, the rules of ordinary multiplication are:

$$0 \times 0 = 0; 0 \times 1 = 0; 1 \times 0 = 0; 1 \times 1 = 1.$$

They too can be summarised in a similar table of arithmetical multiplication:

Arithmetical Multiplication

×	0	1
0	0	0
1	0	1

A glance at the tables of logical and arithmetical addition shows that they are identical. So also are those of logical and arithmetical multiplication.

Now the ideal calculating machine must be such that with an initial input of data it turns out the final answer with as little human interference as possible until the very end. This means that after the initial insertion of the numerical data the machine must not only be able to perform the computation but also be able to decide between the various contingencies that may arise during the course of the calculation in the light of *instructions* also inserted into it along with the numerical data at the beginning. In other words, a calculating machine must also be a logical machine capable of making a choice between 'yes' and 'no', the choice of adopting one or other of two alternative courses open to it at each contingency that may arise during the course of the computation. It is because of the formal identity of the rules of the logical and arithmetic calculi (in the binary notation) that the apparatus designed to mechanise calculation is also able to mechanise processes of thought. That is why the binary system is superior to other systems in both arithmetic and logic.

Another advantage of the binary system is this. A calculating machine can operate in only two ways. First, it may consist of a device which translates numbers into physical quantities measured on specified continuous scales—such as lengths, angular rotations, voltages, *etc.* After operating with these quantities it measures some physical magnitude which gives the result.* Second, it may consist of a device which operates with numbers directly in their digital form by counting discrete objects such as the teeth of a gear-wheel, or discrete events such as electrical pulses. Such, for instance, is the case with the ordinary desk calculating machines like the Brunsviga and Marchant. Naturally, the accuracy of the first type

* For example, a product xy may be evaluated by converting the logarithm of numbers x,y into lengths on a slide rule. We first read the length corresponding to the logarithm of the number x and add to it the length corresponding to that of number y. We then read the number corresponding to the combined lengths to obtain the product xy.

depends on the accuracy of the construction of the continuous scale, and that of the second on the sharpness with which the discrete set of events, such as wheel teeth or electrical pulses, can be distinguished from one another. Since it is easier to distinguish a set of discrete events than to construct a fine continuous scale, the latter type, *viz.* the digital machine, is preferable for highly accurate work. Further, since it is easier to distinguish between two discrete events than ten, digital machines constructed on the binary scale are superior to those on the decimal scale. In other words, the structure of the ideal machine should be a bank of relays each capable of two conditions—say, 'on' and 'off'; at each stage the relays must be able to assume positions determined by the position of some or all of the relays of the bank at a previous stage. This means that the machine must incorporate a clocking arrangement for progressing the various stages by means of one or more central clocks.

Now, as Norbert Wiener has remarked, the human and animal nervous systems, which too are capable of the work of a computation system, contain elements—the nerve cells or neurons—which are ideally suited to act as relays:

'While they show rather complicated properties under the influence of electrical currents, in their ordinary physiological action they conform very nearly to the "all-or-none" principle; that is, they are either at rest, or when they "fire" they go through a series of changes almost independent of the nature and intensity of the stimulus.' This fact provides the link between the art of calculation and the new science of Cybernetics, recently created by Norbert Wiener and his collaborators.

This science (cybernetics) is the study of the 'mechanism of control and communication in the animal and the machine', and bids fair to inaugurate a new social revolution likely to be quite as profound as the earlier Industrial Revolution inaugurated by the invention of the steam engine. While the steam engine devalued brawn, cybernetics may well devalue brain—at least, brain of a certain sort. For the new science is already creating machines that imitate certain processes of thought and do some kinds of mental work with a speed, skill and accuracy far beyond the capacity of any living human being.

The mechanism of control and communication between the brain and various parts of an animal is not yet clearly understood. We still do not know very much about the physical process of thinking in the animal brain, but we do know that the passage of some kind of physico-chemical impulse through the nerve-fibres between the nuclei of the nerve cells accompanies all thinking, feeling, seeing, *etc*. Can we reproduce these processes by artificial means? Not exactly, but it has been found possible to imitate them in a rudimentary manner by substituting wire for nerve-fibre, hardware for flesh, and electro-magnetic waves for the unknown impulse in the living nerve-fibre. For example, the process whereby flatworms exhibit negative phototropism—that is, a tendency to avoid light—has been imitated by means of a combination of photocells, a Wheatstone bridge and certain devices to give an adequate phototropic control for a little boat. No doubt it is impossible to build this apparatus on the scale of the flatworm, but this is only a particular case of the general rule that the artificial imitations of living mechanisms tend to be much more lavish in the use of space than their prototypes. But they more than make up for this extravagance by being enormously faster. For this reason, rudimentary as these artificial reproductions of cerebral processes still are, the thinking machines already produced achieve their respective purposes for which they are designed incomparably better than any human brain.

As the study of cybernetics advances—and it must be remembered that this science is just an infant barely ten years old—there is hardly any limit to what these thinking-machines may do for man. Already the technical means exist for producing automatic typists, stenographers, multilingual interpreters, librarians, psychopaths, traffic regulators, factory-planners, logical truth calculators, *etc.* For instance, if you had to plan a production schedule for your factory, you would need only to put into a machine a description of the orders to be executed, and it would do the rest. It would know how much raw material is necessary and what equipment and labour are required to produce it. It would then turn out the best possible production schedule showing who should do what and when.

Or again, if you were a logician concerned with evaluating the logical truth of certain propositions deducible from a set of given premises, a thinking machine like the Kalin-Burkhart Logical Truth Calculator could work it out for you very much faster and with much less risk of error than any human being. Before long we may have mechanical devices capable of doing almost anything from solving equations to factory planning. Nevertheless, no machine can create more thought than is put into it in the form of the initial instructions. In this respect it is very definitely limited by a sort of conservation law, the law of conservation of thought or instruction. For none of these machines is capable of thinking anything new.

A 'thinking machine' merely *works out* what has already been thought of beforehand by the designer and supplied to it in the form of instructions. In fact, it obeys these instructions as literally as the unfortunate Casabianca boy, who remained on the burning deck because his father had told him to do so. For instance, if in the course of a computation the machine requires the quotient of two numbers of which the divisor happens to be zero, it will go on, Sisyphus-wise, trying to divide by zero for ever unless expressly forbidden by prior instruction. A human computer would certainly not go on dividing by zero, whatever else he might do. The limitation imposed by the aforementioned conservation law has made it necessary to bear in mind what Hartree has called the 'machine-eye view' in designing such machines. In other words, it is necessary to think out in advance every possible contingency that might arise in the course of the work and give the machine appropriate instructions for each case, because the machine will not deviate one whit from what the 'Moving Finger' of prior instructions has already decreed. Although the limitation imposed by this conservation law on the power of machines to produce original thinking is probably destined to remain for ever, writers have never ceased to speculate on the danger to man from robot machines of his own creation. This, for example, is the moral of stories as old as those of Famulus and Frankenstein, and as recent as those of Karel Čapek's play, *R.U.R.*, Olaf Stapledon's *First and Last Men.*

It is true that as yet there is no possibility whatsoever of constructing Frankenstein monsters, Rossum robots or Great Brains—that is, artificial beings possessed of a 'free will' of their own. This, however, does not mean that the new developments in this field are without danger to mankind. The danger from the robot machines is not technical but social. It is not that they will disobey man but that if introduced on a large enough scale, they are liable to lead to widespread unemployment.

3

THF CALCULUS

THE knowledge that things, in spite of their apparent permanence, are really in a state of perpetual flux and change, is probably as old as human civilisation. This knowledge formed the basis of philosophical speculations about flux and change which preceded the mathematical formulations by whole millennia. Long before Greek civilisation, the mystical view of change, that nothing really exists but only flux or flow—a view revived more recently in Bergson's *Creative Evolution*—had evolved from commonplace observations. But the mathematics (as opposed to the mysticism of flux) originated during the second half of the seventeenth century, when mathematicians began to study the problem of change.

Speaking of the mathematics of flux, Leibnitz, who shares with Newton the honour of inventing it, said, 'My new calculus . . . offers truth by a kind of analysis and without any effort of the imagination—which often succeeds only by accident; and it gives us all the advantages over Archimedes that Vieta and Descartes have given us over Apollonius.' This claim was no exaggeration, for the calculus proved to be the master key to the entire technological progress of the following three centuries. What then is this calculus of Newton and Leibnitz which has had such momentous consequences?

Mathematically, the calculus is designed to deal with the fundamental problem of change, *viz*. the rate at which anything changes or grows. You cannot even begin to answer questions like this unless you know what it is that changes and how it changes. What you need, in fact, is a growth function—that is, a correlation of its growth against the flow of time. But growth and flow are rather vague terms and must first be given precise mathematical expression. Take first the flow of time. Our daily perception of events around us, not to speak of the physiology of our own bodies, makes us aware of what we call the flow of time. This means that every one of us can arrange the events that we perceive in an orderly sequence. In other words, we can tell which of any two events perceived occurred earlier and which later. By means of physical appliances, such as a watch, we can even say how much earlier or later. This enables us to particularise or 'date' the events by a number—the number of suns, moons or seconds that have elapsed since a certain beginning of time. This is the commonsense way of reckoning time, which both Newton and Leibnitz took for granted.

Now about growth. Here again, to fix ideas we may think of something concrete that grows, *e.g.* the weight of a newly born baby. We could say something about its growth if we weighed it on a number of different occasions. If we 'dated' the occasions we should get a succession of number pairs, one number denoting the date of weighing and the other the corresponding weight on that date. This table of number pairs, *i.e.* the weight and its corresponding date, is, in fact, the growth function—the mathematical representation of what we have called the correlation of growth against the flow of time.

We must first somehow infer the growth function of changing phenomena before the calculus can be applied, and our guide in this matter is mostly experience. Historically, the first phenomenon to be studied by the calculus was the motion of material bodies, such as that of a stone rolling down a hillside. Galileo inferred the growth function of rolling stones from studying its laboratory replica. He allowed balls to roll down inclined planes and observed the distances travelled by them at various times or 'dates' such as at 1, 2, 3, . . . minutes after the commencement of the roll. If we repeat his experiment we may, for example, find the following table of corresponding number pairs:

Time (t) in seconds:	0, 1, 2, 3, 4, 5, . . .
Distance rolled (y) in feet:	0, 1, 4, 9, 16, 25, . . .

It is clear that we may replace this table by the formula $y = t^2$, which is called the growth function of the distance. The problem that the calculus is designed to handle is as follows. Knowing the growth function of the distance the ball rolls, can we discover how fast it moves? In other words, what is its speed?

It is obvious that the ball does not travel with the same speed during the whole course of its motion. The speed itself continually grows. The calculus seeks to derive the growth function of the speed, knowing that of the distance, and thus to give us a formula which enables us to obtain the speed at any given time. Suppose we wish to find its speed at the time denoted by $t = 2$. Take any other time close to $t = 2$ but slightly later, say $t = 2{\cdot}1$. The distances travelled in these two times are $(2)^2$ and $(2{\cdot}1)^2$ units of distance (say feet), respectively, for the distance rule, $y = t^2$, applies equally when t is a fraction. The distance travelled during the time interval from $t = 2$ to $t = 2{\cdot}1$ is therefore $(2{\cdot}1)^2 - (2)^2$. Hence, the *average* speed during the interval of time from 2 to 2·1, or 0·1 units of time (say seconds) is $\dfrac{(2{\cdot}1)^2 - (2)^2}{0{\cdot}1} = 4{\cdot}1$ feet per second. We could regard it as the *actual* speed of the ball at the time $t = 2$ for all ordinary purposes; but if someone insisted that it was only an average during the interval taken and not the actual speed at $t = 2$, we should have to concede the point. To satisfy our opponent, we might calculate the average speed during still shorter intervals of time, say, that during the intervals (2, 2·01), (2, 2·001), (2, 2·0001), . . ., ... In exactly the same way as before, we find that the respective averages during these shorter intervals are 4·01, 4·001, 4·0001, . . ., . . . feet per second. This shows that the average continually approaches 4 as we successively shorten the interval used for averaging. More precisely, we can make it come as close to 4 as we like provided only we take a sufficiently small interval of time for calculating the average. This value 4 is then the *limit* of the average—it is a sort of terminal point or bound below which the average can never fall. We are, therefore, justified in taking this limit as the actual speed at $t = 2$. If we try to work out the speed at other times such as $t = 1, 3, 4, 5, \ldots$ in the same way, we find it to be 2, 6, 8, 10, . . . respectively. Thus the table of corresponding values of speed at various times is:

Time (t) in seconds:	1, 2, 3, 4, 5, . . .
Speed (v) in feet per second:	2, 4, 6, 8, 10, . . .

This suggests the formula $v = 2t$ for the growth function of speed.

We could treat the speed function in exactly the same way as we have treated the distance function. In other words, we could now enquire how fast does the speed change or what is the acceleration of the ball. We proceed in the same manner as before by taking an instant of time, say 2·1, very close to the instant 2. The speed changes from 2(2) to 2(2·1) so that the average increase in speed—or the acceleration—is

$$\frac{2(2{\cdot}1) - 2(2)}{(2{\cdot}1 - 2)} = 2.$$

We get the same result no matter how close to 2 we choose the second instant of time for averaging the increase in speed. We may, therefore, regard this average acceleration as the actual acceleration at $t = 2$. If we try to work out in the same manner the acceleration at other instants of time, such as $t = 1, 3, 4, \ldots$, we find that it remains 2 all the time. Consequently, the table of corresponding values of acceleration at various times is:

Time (t) in seconds:	0, 1, 2, 3, 4, . . .
Acceleration (a):	2, 2, 2, 2, 2, . . .

This suggests the formula $a = 2$, which means that acceleration remains 2 at all instants of time. In other words, the ball rolls with uniform acceleration.

Although the calculus was originally devised to calculate time rates of change like speeds and accelerations of moving bodies. its technique is equally applicable to all sorts of rates of change. In everyday life we usually come across time rates, such as interest, speed, acceleration, growth, *etc.*, but whenever we have one variable quantity, y, depending on another variable, x, we may enquire about the rate of change of y per unit change of x. Thus the irrigation engineer is interested in the hydraulic pressure, y, that the dam surface has to endure at any given depth, x, below the water surface. Here, although the water pressure is assumed to remain static everywhere, we may, nevertheless, legitimately enquire how fast the water pressure rises with increasing depth. In other words, we may wish to determine the rate of change of pressure per unit change of depth.

Whenever we have a pair of magnitudes, y and x (such as hydraulic pressure and depth, freight or fare payable and distance of journey, income tax and income, *etc.*), so related that the measure of one depends on that of the other, the former is said to be a function of the latter. Symbolically, we denote this dependence by the expression $y = f(x)$, where f is only a shorthand for 'depends on'.

Now since dependence like so many other relations, such as friendship, is a reciprocal relation, we could equally regard the measure of x as depending on that of y. Thus, if y is a function of x, then equally x is some function F of y. The function $F(y)$ is called the inverse of $f(x)$. For instance, if $y = 3x - 6$, then $x = y/3 + 2$. Hence the inverse of y, $y = f(x) = 3x - 6$ is the same as $x = F(y) = y/3 + 2$. Although, strictly speaking, either variable may be expressed as a function of the other, in most situations it is more natural to regard the variation of one as independent of and, in a way, 'controlling' that of the other. For instance, in the above-mentioned illustrations it is more natural to consider income tax as a function of income, railway freight as a function of distance, or hydraulic pressure as a function of depth, rather than the other way about. Income tax, railway freight and hydraulic pressure are therefore dependent variables, as they do, in a real way, depend on income, distance and depth, respectively. It is true that there are some cases in which it may not be quite obvious which of the two variables should be treated as independent, but in such cases we may make any choice to suit our convenience.

In most cases the dependence between the two variables y, x is far too complicated to be reduced to a formula. For example, the bitterness of Swift's satire or the pungency of Carlyle's invective may, perhaps, if they could be measured, be functions of the amount of bile secreted by their livers at the times of writing. But if so, no formula can be devised to express this dependence. In science, however, we mostly deal with functional dependences which can be reduced to a formula. For a closer peep into the working of nature, where we also come across dependences far too complex to be trapped in the neat expression of a single formula, mathematicians during the last century or so have been obliged to consider functional dependences, which may be expressed by a series of formulae instead of by a single formula. As an instance of such a function we may cite the relation between pressure, y, and volume, x, of a gas at constant temperature. For ordinary pressures such as could be applied in the time of Boyle, the formula, $y = constant/x$, usually known as Boyle's law, expresses this dependence. For other ranges of pressure such as are now possible, different formulae have to be used. The functional relation between pressure and volume of a gas is thus a multi-formula function. The same is true of most laws of nature.

Now, whatever y and x may be, and whatever the formula or formulae expressing their dependence, we may want to know the rate of change of one per unit increase of the other. Suppose $y = f(x)$ is a functional relation between any dependent variable y and an independent variable x. Let us calculate the rate of change of $f(x)$ per unit change of x. In general, this rate will itself vary and depend on the particular value of x chosen. Suppose we want it for $x = 2$, then, as in the case of time when we were discussing speed above, we choose a short range of the independent variable lying between 2 and $2 + h$, h being a small number. If the independent variable varies from 2 to $2 + h$, the dependent variable y will change from $f(2)$ to $f(2 + h)$. The net change of its value over the range $(2, 2 + h)$ is $f(2 + h) - f(2)$. The average rate of change over this range is, therefore,

$$\frac{f(2 + h) - f(2)}{h}.$$

If h is reasonably small, we may take this *average* rate as the actual rate of change of $f(x)$ at $x = 2$. But if someone insisted that this is only an average and not the actual rate, we should have to accept the validity of his objection. To avoid such an objection, we say that the actual rate of change of y at $x = 2$ is the *limit* of the average rate of change of y per unit change of x, when the interval of x, over which it is averaged, is decreased indefinitely. The idea is exactly analogous to that of speed, with x substituted for t. This limit is known as the differential co-efficient or derivative of y with respect to x at $x = 2$, and is usually written as $\left(\frac{dy}{dx}\right)_{x=2}$, or $f'(2)$.

We could do to $f'(x)$ what we did to $f(x)$, that is, we could enquire what is the rate of change of $f'(x)$ with respect to x at the value $x = 2$—or at any other value, for that matter. It is similarly the limit of the average rate of change:

$$\frac{f'(2 + h) - f'(2)}{h} \quad . \quad . \quad . \quad . \quad (1)$$

as h tends to zero. This limit is the differential co-efficient or derivative of $f'(x)$ with respect to x at $x = 2$. Since $f'(x)$ itself was derived from $f(x)$ by the same process, we may also call the limiting value of (1) the *second* differential coefficient or derivative of $f(x)$ with respect to x at $x = 2$. We denote it by $\left(\frac{d^2y}{dx^2}\right)_{x=2}$, or $f''(2)$. There is no reason why we should stop here. We could repeat the process to derive successively the third, fourth . . . differential coefficients of $f(x)$ for any specific value of x, such as 2. This successive generation of differential coefficients is no mere free

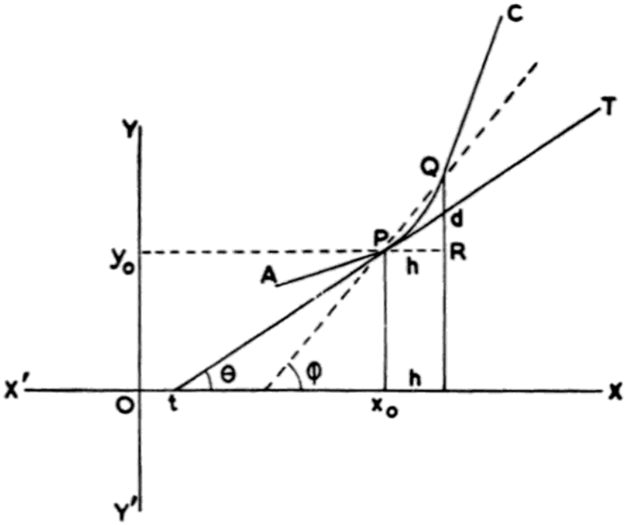

FIG. 2—As Q travels along the graph line to coincide with P, PQ takes the position of the tangent PT. The angle ϕ becomes the angle θ. In the limit therefore, $\frac{dy}{dx} = \tan \theta$.

creation of the curious mind. The railroad engineer has to employ second derivatives to calculate the curvature of the line he constructs. He needs a precise measure of the curvature to find the exact degree of banking required to prevent trains from overturning. The automobile designer utilises the third derivative in order to test the riding quality of the car he designs, and the structural engineer has even to go to the fourth derivative in order to measure the elasticity of beams and the strength of columns.

It is not difficult to see how the first and higher order derivatives arise naturally in problems like these. Suppose the curve $APQC$ represents the railway line. (See Fig. 2.) Before the calculus can be applied we must somehow represent it algebraically, that is, by some functional formula. This is done by taking any two perpendicular reference lines XOX' and

YOY' through a point of origin O and indicating the position of any point P by its distances (x, y) from these reference lines or 'axes'. The distances, x, y, are called co-ordinates. Now if we consider any point P on the line and measure its co-ordinates (x, y), we find that y is usually some function of x, such as $y = f(x)$.

This equation gives the variation of y due to a change of x. At any point $P(x_0, y_0)$ of the line the value of y_0 is $f(x_0)$. At another point $Q(x_0 + h, y_0 + d)$ very close to P and also on the railway line, the value of y is $f(x_0 + h)$. Hence the change in y is $f(x_0 + h) - f(x_0)$, or the length $QR = d$. The average rate of change of y per unit change of x is

$$\frac{f(x_0 + h) - f(x_0)}{h} = \frac{d}{h} = \frac{QR}{PR}.$$

If the angle QPR is denoted by φ, the ratio QR/PR is known as the *slope* or gradient of the line PQ, and is written as $\tan \varphi$, which is short for tangent of the angle φ. It follows, therefore, that the average rate of change of y at P with respect to x is the gradient or slope of the chord joining P to another point Q of the line very close to P. If we diminish h, the average rate of change of y during the interval $(x_0, x_0 + h)$ approximates more and more closely to the actual value of the instantaneous rate at P. The latter has already been defined as dy/dx. On the other hand, as h diminishes and Q approaches P along the graph line, the chord line PQ becomes a tangent to it at P, that is, the straight line PT which just grazes it. Hence the value of dy/dx at P measures the gradient or slope of the tangent to the curve at P.

$$\text{In symbols, } \frac{dy}{dx} = \tan \theta.$$

Now, at any point P the direction of the line is along the tangent PT. At another point Q on the line it is along the tangent QT' at Q. (See Fig. 3.) The measure of its bend as we travel from P to Q along the line is therefore the angle between the two tangents at P and Q. The rate of its bend or curvature is the ratio:

$$\frac{\text{total bend}}{\text{length of the curved line } PQ}.$$

This is, of course, the average curvature along the whole length of the arc PQ. If we want its precise curvature *at* the point P, we must, as before, find the limit of this ratio as the length of the arc PQ is indefinitely decreased by bringing Q infinitely close to P. In other words, we have to differentiate the angle θ of the tangential direction at P with respect to the length PQ of the arc. But, as we have seen already, $\tan \theta$ is dy/dx. It is,

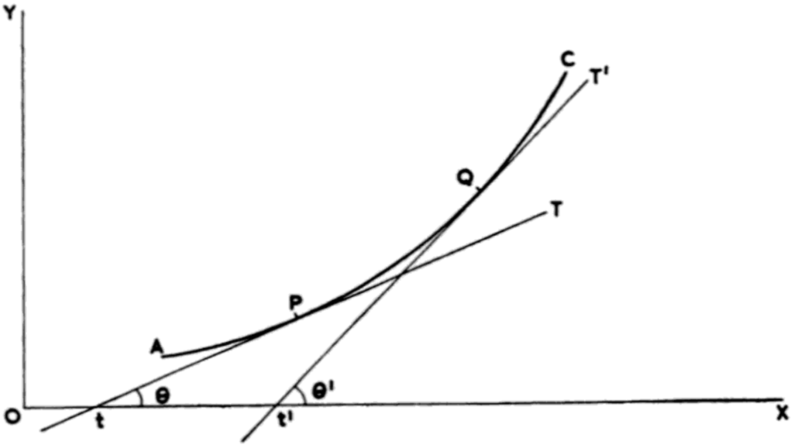

FIG. 3—The total bend of the line as we move from P to Q is the angle between the two tangents at P and Q, that is, $\theta' - \theta$.

therefore, inevitable that in the process of differentiation of the angle θ we should encounter second derivatives.

* * * *

As we saw, the problem of calculating instantaneous speed and acceleration of moving bodies *at* any given instant of time from the mathematical formula connecting the distance (y) travelled and the time (t) of their fall or flight, gave rise to the differential calculus. The inverse problem of calculating the distance travelled, given a mathematical formula connecting speed or acceleration with time, led to the development of the *Integral Calculus.*

Suppose a particle moves with speed v which is given by the relation $v = t$. What is the distance, s, travelled during the time interval from say $t = 1$ to $t = 6$? If the speed had been uniform we could have got the desired distance by multiplying the uniform speed by the duration of the time interval during which it was maintained. But in the problem before us the speed does not remain the same even for two consecutive instants. How are we to apply the rule which is valid for a static speed to calculating the distance travelled when it is no longer so?

Essentially this is a problem of reconciling irreconcilables, of finding a method of resolving the inherent conflict between change and permanence. This conflict is universal—between the ever-changing material world and the static, permanent or quasi-permanent forms and categories that we invent and impose upon the world to understand it. Newton and Leibnitz were the first to devise a practical way of resolving this conflict in the case of moving bodies. The method proposed was simple in principle—once it was discovered. They divided the duration of motion into a large number of sub-intervals during each of which its velocity was assumed *not* to change. The rule for static speed could now be applied for each sub-interval *separately* and the distance derived by summing all the distances travelled in each sub-interval.

Suppose, for instance, we divide the interval of time from $t = 1$ to $t = 6$ into any number, say 10, of equal sub-intervals, each of duration $\frac{6-1}{10} = \frac{1}{2}$ second, by taking nine intermediate point-instants $1 + \frac{1}{2}$, $1 + \frac{2}{2}$, $1 + \frac{3}{2}$, $1 + \frac{4}{2}$, . . ., $1 + \frac{9}{2}$ between the times $t = 1$ and $t = 6$. Consider now the first sub-interval from the initial instant $t = 1$ to $t = 1 + \frac{1}{2}$. Since the speed v is given by the formula $v = t$ the speeds at the beginning and end of the first sub-interval were 1 and $(1 + \frac{1}{2})$ respectively. The distance travelled during the first sub-interval is, therefore, greater than $1(\frac{1}{2})$ but less than $(1 + \frac{1}{2})\frac{1}{2}$. Similarly the limits between which the distances travelled during the second, third, fourth, . . . and tenth sub-intervals lie, can be calculated. We tabulate these limits below:

Sub-interval of time	*Lower limit of the distance travelled*	*Upper limit of the distance travelled*
First	$1(\frac{1}{2})$	$(1 + \frac{1}{2})\frac{1}{2}$
Second	$(1 + \frac{1}{2})\frac{1}{2}$	$(1 + \frac{2}{2})(\frac{1}{2})$
Third	$(1 + \frac{2}{2})(\frac{1}{2})$	$(1 + \frac{3}{2})(\frac{1}{2})$
Fourth	$(1 + \frac{3}{2})(\frac{1}{2})$	$(1 + \frac{4}{2})(\frac{1}{2})$
Fifth	$(1 + \frac{4}{2})(\frac{1}{2})$	$(1 + \frac{5}{2})(\frac{1}{2})$
Sixth	$(1 + \frac{5}{2})(\frac{1}{2})$	$(1 + \frac{6}{2})(\frac{1}{2})$
Seventh	$(1 + \frac{6}{2})(\frac{1}{2})$	$(1 + \frac{7}{2})(\frac{1}{2})$
Eighth	$(1 + \frac{7}{2})(\frac{1}{2})$	$(1 + \frac{8}{2})(\frac{1}{2})$
Ninth	$(1 + \frac{8}{2})(\frac{1}{2})$	$(1 + \frac{9}{2})(\frac{1}{2})$
Tenth	$(1 + \frac{9}{2})(\frac{1}{2})$	$(1 + \frac{10}{2})(\frac{1}{2})$
Total =	s_1	S_1

Let s_1 and S_1 be the sums of the lower and upper limits of distances travelled during the ten sub-intervals. If we add up the ten terms shown in the above table of lower and upper limits we get,

$$s_1 = \tfrac{1}{2}\{10 + (1 + 2 + 3 + \ldots 9)\tfrac{1}{2}\}$$
$$= \tfrac{1}{2}(10 + \tfrac{45}{2}) = 16{\cdot}25;$$

$$S_1 = \tfrac{1}{2}\{10 + (1 + 2 + 3 + \ldots 10)\tfrac{1}{2}\}$$
$$= \tfrac{1}{2}(10 + \tfrac{55}{2}) = 18{\cdot}75.$$

The actual distance travelled, therefore, lies between 16·25 and 18·75 feet. It is not an exact answer but good enough for rough purposes. To improve the precision of our answer, we have to shorten further the sub-intervals by dividing the original interval $t = 1$ to $t = 6$ into a larger number of sub-intervals, say, 100 instead of 10 so that the duration of each sub-interval now is $\frac{6-1}{100} = \frac{1}{20}$. The only snag is that the calculation is longer but not different in principle. If we denote by s_2 and S_2 the corresponding sums of lower and upper limits for the distance travelled during these shorter sub-intervals, we find that

$$s_2 = \tfrac{1}{20}\{100 + (1 + 2 + 3 + \ldots\ 99)\tfrac{1}{20}\} = 17{\cdot}37;$$

and, $$S_2 = \tfrac{1}{20}\{100 + (1 + 2 + 3 + \ldots\ 100)\tfrac{1}{20}\} = 17{\cdot}63.$$

This gives us still closer limits within which the actual distance travelled must lie, *viz.* 17·37 and 17·63. If we repeat the calculation by dividing the interval into 1000 sub-intervals, the corresponding sums s_3 and S_3 will be found to be

$$s_3 = \tfrac{1}{200}\{1000 + (1 + 2 + 3 + \ldots\ 999)\tfrac{1}{200}\} = 17{\cdot}48;$$

and, $$S_3 = \tfrac{1}{200}\{1000 + (1 + 2 + 3 + \ldots\ 1000)\tfrac{1}{200}\} = 17{\cdot}51.$$

These are even closer limits for the actual distance travelled during the interval (1, 6). We may tabulate the successive values of the upper and lower limits:

$s_1 = 16{\cdot}250$	$S_1 = 18{\cdot}750$
$s_2 = 17{\cdot}375$	$S_2 = 17{\cdot}625$
$s_3 = 17{\cdot}4875$	$S_3 = 17{\cdot}5125$
$s_4 = 17{\cdot}49875$	$S_4 = 17{\cdot}50125.$

From the above table, we may infer that the sums of lower limits or, for short, lower sums $s_1, s_2, s_3, \ldots$ continually increase while the upper sums $S_1, S_2, S_3, \ldots$ continually decrease. Moreover, the difference between the two corresponding lower and upper sums continually decreases so that both sums approach the same limit as the number of sub-intervals is indefinitely increased. This common limit of the two sums which, as we may guess, is 17·5, is the actual distance travelled. It is also known as the integral of speed with respect to time over the interval (1, 6). We denote it by the symbol

$$\int_1^6 v\,dt, \text{ or } \int_1^6 t\,dt, \text{ as } v = t.$$

In this, the integral sign $\int$ is only a distorted form of S, short for sum, to remind us that it is really a limiting sum.

If we look at the speed-time graph (Fig. 4), the line ODC will represent the relation $s = t$. We can use this graph to simplify the calculation of the speed integral. Suppose we divide the time interval (1, 6) represented by the segment *AB* of the time-axis *Ot* into any number of equal sub-intervals. Consider now any such sub-interval, say, *LM*. At the instant of time represented by the point *L*, the speed is given by *LP* and at *M* by *MQ*. The distance actually travelled during the sub-interval *LM*, therefore, lies between the lower limit, *LM.LP*, and the upper limit, *LM.MQ*, since distance

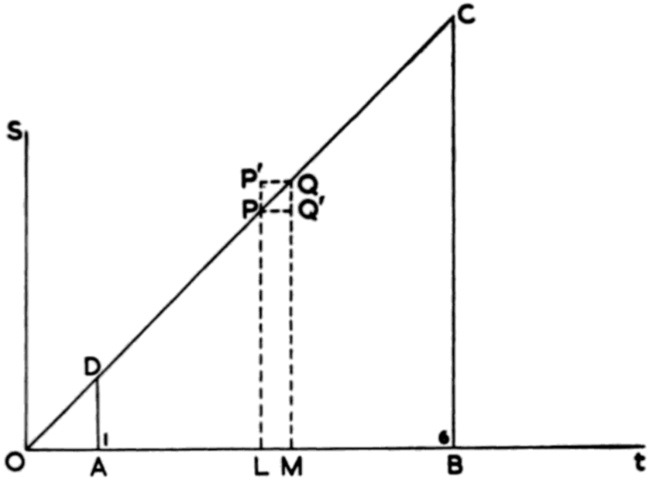

FIG. 4

= time × speed. These limits are obviously the areas of the two rectangles, *LPQ′M* and *LP′QM*. As we increase the number of sub-intervals, the sub-interval *LM* diminishes indefinitely. The areas of the two rectangles between which the actual distance travelled lies approximate more and more closely to the area of the trapezium *LPQM*. But as the total distance travelled during the interval (1, 6) is the sum of all such areas, it is naturally represented by the area of the trapezium *ABCD*. Now the area of the trapezium *ABCD*—

$$= \triangle\, OBC - \triangle\, OAD$$
$$= \tfrac{1}{2}(OB)(BC) - \tfrac{1}{2}(OA)(AD)$$
$$= \tfrac{1}{2}(6)(6) - \tfrac{1}{2}(1)(1)$$
$$= \tfrac{1}{2}(6^2 - 1^2) = 17{\cdot}5.$$

We may therefore write $\int_1^6 t\,dt = \frac{1}{2}(6^2 - 1^2)$.

From this we may readily infer that the distance travelled during *any* interval (t_0, t_1) is the *definite integral* $\int_{t_0}^{t_1} t\,dt = \frac{1}{2}(t_1^2 - t_0^2)$.

We can further generalise this result. Instead of working with the speed-time graph $s = t$, we may start with the graph of any function $y = f(x)$, and enquire what is the area enclosed by the graph line *BC*, the two ordinate lines *AB*, *DC* and the segment *AD* on the *x*-axis, where *OA* is any length x_0 and *OD* any other length x_1. (See Fig. 5.) As before, we divide the segment *AD* into any number of equal sub-intervals. Take any sub-interval *LM*. The area *LPQM* enclosed by the arc *PQ* of the curve evidently lies between the areas of the inner rectangle *LPQ'M* and the outer rectangle *LP'QM*. As we increase the number of these rectangles indefinitely, the areas of the two rectangles approximate more and more closely to the area *LPQM* enclosed by the small arc *PQ* of the graph line. But the sum of the areas of inner rectangles like *LPQ'M* is the exact

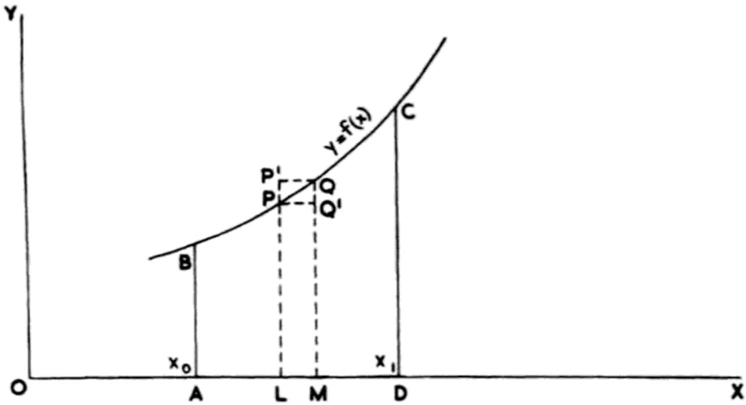

FIG. 5

analogue of the lower sum, and that of the outer rectangles like *LP'QM* that of the upper sum we constructed earlier while integrating the speed graph $v = t$. These two upper and lower sums tend to a common limit as the number of sub-intervals is indefinitely increased. This common limit is the integral $\int_{x_0}^{x_1} f(x)dx$. Obviously, this integral is also the measure of the area *ABCD* enclosed by the graph line *BC* as each of the inner and outer rectangles approximates, more and more closely, to the actual area *LPQM* under an elementary arc *PQ* of the graph line.

Now although we have defined the definite integral $\int_{x_0}^{x_1} f(x)dx$ as the area of the curve *ABCD*, it is not the only possible interpretation that can be given to it. The essential idea behind it is that it is a *limiting sum* of an infinite series of terms. It thus comes to pass that whenever we have to add up an infinite number of values of a function corresponding to an infinite number of values of its independent variable, the definite integral plays an indispensable role. For instance, when the hydraulic engineer constructing a dam wants to ascertain the total water pressure likely to be exerted on the whole face of the dam, he divides the entire dam surface into an infinite number of point-bits. As the pressure at any arbitrary point-bit of the dam surface can be calculated, the calculation of the whole pressure on the dam surface is a summation problem and therefore amenable to integration.

Here pressure is a function of the depth of the point-bit below the water surface. It is the same with other problems facing the bridge engineer, the architect and the electrician in the calculations, respectively, of moments of inertia, centres of gravity of solids and surfaces, magnitudes of electromagnetic fields, *etc.* In all these problems the value of a quantity such as mass, moment, hydraulic pressure, electric or magnetic forces, is given at each of an infinite number of points in a region or space and it is desired to calculate their sum.

It is often inconvenient to calculate such limiting sums directly. The calculation is greatly simplified by the use of a theorem which links integration with differentiation. As we have seen, the integral $\int_{x_0}^{x_1} f(x)dx$ is the area enclosed by *ADCB* (Fig. 5). If we treat x_1 in this integral as a variable, the integral itself becomes a function of x_1. Let us call it $F(x_1)$. Then

$$F(x_1) = \int_{x_0}^{x_1} f(x)dx.$$

What is the differential coefficient of $F(x_1)$? According to the rule we have established, it is the limit of

$$\frac{F(x_1 + h) - F(x_1)}{h}$$

as h tends to zero. Now $F(x_1)$ is the area *AQPB* (Fig. 6) and $F(x_1 + h)$ is the area *AQ′P′B* (Fig. 6). The difference is the area *QQ′P′P* or approximately $QQ'.PQ = h.f(x_1)$. The smaller h is, the more nearly is the area *PQQ′P′* equal to $h.f(x_1)$ and hence as h tends to zero, the limiting value of

$$\frac{F(x_1 + h) - F(x_1)}{h} = \frac{h.f(x_1)}{h} = f(x_1), \text{ is } f(x_1).$$

In other words, $\dfrac{dF}{dx_1} = \dfrac{d}{dx_1}\displaystyle\int_{x_0}^{x_1} f(x)dx = f(x_1).$

That is, the differential coefficient of the integral of a function is the function itself. Thus differentiation and integration are inverse operations like multiplication and division. Just as multiplying a number by another and

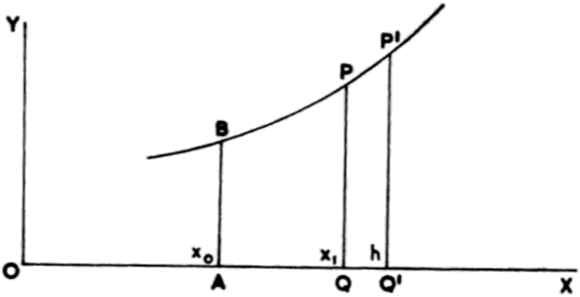

FIG. 6—$F(x_1)$ is the area $AQPB$, and $F(x_1 + h)$ the area $AQ'P'B$. The difference, therefore, is the area $QQ'P'P$ which is approximately $QQ'.PQ$, or $f(x_1).h$. The rate of change of area under the curve at P is the value of $f(x)$ at P.

then dividing the product by the same factor leaves the number unchanged, so integrating a function $f(x)$ and then differentiating the integral leaves the function as it was before.

The great power of the calculus depends on this fundamental theorem, for, in our study of nature we often assume a formula for the rate of growth

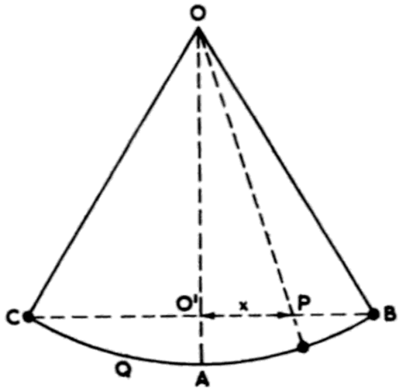

FIG. 7

of one variable per unit change of the independent variable, and we then want to know if the formula really works. For instance, in his study of motion Newton assumed that the temporal rate of change of velocity of a moving body, that is, its acceleration, is equal to F/m, where F is the force applied to it and m is its mass. He applied this formula to a wide class of phenomena, from the motion of the pendulum to that of the moon.

This formula always led to an equation in which occurred the rates of change or the differential coefficients. Thus, in the case of the motion of the pendulum bob, it led to the equation $dv/dt = F/m$.

Now, if we assume that the circular arc that the bob describes is small, we may consider it as oscillating along the straight line BC instead of the arc BAC (see Fig. 7). Let its distance in any position P be x from the central position O'. The force F acting on it in this position is that part or component* of the tension of the string holding the bob, which acts along the direction of its motion, that is, the direction $O'B$. It can be shown that this component is $-mg.x/l$, where l is the length of the string and mg the weight† of the bob. Newton's formula, therefore, leads to the equation

$$dv/dt = -gx/l.$$

But as we know, v itself is dx/dt, so that the equation of the motion of the pendulum bob is

$$\frac{dv}{dt} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2} = -\frac{gx}{l}$$

or,

$$d^2x/dt^2 = -gx/l \quad . \quad . \quad . \quad . \quad (1)$$

Such an equation in which differential coefficients of the dependent variable with respect to the independent variable occur is known as a differential equation. All that Newton's famous laws of motion and gravitation do is to set up a system of one or more differential equations essentially of the type written above. The motion of the body or system of bodies is then known, if we can 'solve' the differential equations, that is, find x as a function of t so that it satisfies them.

If we consider the solution of any differential equation such as (1) we shall find that it is indeterminate in one respect. You will recall that in formulating it, we merely used Newton's law and took no account of the initial position from where the bob is let go. This means that the differential equation is indifferent to its initial position.‡ No matter where the bob is, one and the same differential equation (1) results, and yet what motion it actually executes must also depend on its initial position. Thus its motion when it is released at B is not the same as when it is released at P. In the one case it oscillates between B and C and in the other between P and Q (see Fig. 7). It therefore follows that the differential equation defines no

* See Chapter 4 for a fuller explanation of the component of a force.

† Weight, by the way, is the product of mass m and the acceleration g of gravity, the acceleration with which everything when released falls towards the surface of the earth.

‡ The assumption is still made that the arc BC is small and can be identified with the chord BC.

particular motion of the bob but the whole class of its possible motions corresponding to all its initial positions.

The precise solution of the equation therefore involves the choice of that particular solution out of this class which fits the prescribed initial condition. Here the initial condition is specified by a single magnitude denoting the position of the bob just at the commencement of its motion. But in more complicated cases there are usually many initial conditions which have to be specified by several magnitudes. Take, for instance, Lord Kelvin's theory of the transmission of signals in a submarine cable, whereby he established the theoretical feasibility of an Atlantic cable long before trans-oceanic cablegram became an accomplished fact.

In this case, not only is the differential equation more complicated but the initial conditions too are more numerous. The reason is that we need to know the state of the cable before one of its ends is suddenly connected to a battery terminal, and not merely at one particular point but all along its length. In other words, the initial conditions also include boundary conditions—that is, the initial state of affairs prevailing all over the boundary or surface of the cable.

From what has just been said it follows that even if we are somehow able to solve a differential equation, we must adjust it to suit the given initial conditions. A vast amount of new mathematics created during the past 200 years is merely the outcome of this search in physics and astronomy for solutions of differential equations satisfying prescribed initial conditions. After a prodigious number of special solution functions had been invented, it was found that in many cases a differential equation along with its accompanying set of initial conditions is equivalent to an *integral* equation, in which the unknown variable x appears under an integral sign instead of a sign of derivation. It is thus possible to reformulate certain physical problems in terms of integral rather than differential equations. The great advantage of such reformulation is that the passage from simpler to more difficult problems is not attended with any serious increase in complication as is the case with differential equations.

There are, however, phenomena in which the initial state of a system does not suffice to determine its subsequent evolution. Thus, the evolution of an elastic system is not always determined by its initial state alone but by all its previous states. The past is not completely obliterated and an ancestral influence of a sort controls the shape of things to come. For instance, the type of motion a pendulum bob executes depends solely on the initial position from where it is released no matter what oscillations it might have executed earlier, but if you try to twist an elastic wire, the result will largely depend on how and how many times it had been twisted before. It is because of this hereditary or ancestral influence that materials sometimes fatigue and break down under comparatively minor strains, whereas earlier they could have stood up to much heavier strains.

Such hereditary phenomena require a new mathematical instrument called the integro-differential equation, in which the unknown function x appears under a sign of both integration and derivation. It can be shown that such an equation is equivalent to an infinite number of ordinary differential equations, and this is the reason why they apply to hereditary phenomena. For the initial state which was determined earlier by a few magnitudes has now to be broadened to include the whole infinity of past states. It has, therefore, to be specified by an appropriate choice of an infinite number of magnitudes.

Now, if we view the elastic system atomistically, and thus consider an extremely large number of ordinary differential equations regulating the motions of an extremely large number of particles, each with its own set of initial conditions, we may disregard the hereditary feature which experience seems to require. But as an infinity of differential equations is mathematically intractable, it has to be replaced by a single integro-differential equation which is equivalent to this infinite set.

It is needless to add that it is easier to write equations, whether differential, integral or integro-differential, than to solve them. Only a small class of such equations has been solved explicitly. In some cases, when, due to its importance in physics or elsewhere, we cannot do without an equation of the insoluble variety, we use the equation itself to define the function, just as Prince Charming used the glass slipper to define Cinderella as the girl who could wear it. Very often the artifice works; it suffices to isolate the function from other undesirables in much the same way as the slipper sufficed to distinguish Cinderella from her ugly sisters.

* * * *

(*The following section may be omitted on first reading*)

So far we have considered only functions which depend on a single independent variable x. This assumes that variation in a magnitude (y) can be explained by the variation in a single independent variable x. But in most situations, where we have to explain the variation of a magnitude (y), this assumption is simply not true. As a rule a dependent variable y is influenced by several independent variables simultaneously. Even in the simple case cited earlier, of the pressure of a gas enclosed in a cylinder, it is rather an over-simplification to say that pressure depends on volume, because in actual fact it also depends very materially on temperature too. In this case this extra dependence on temperature does not cause much inconvenience, for we can study the functional relationship between pressure and volume by keeping the temperature constant. But in many cases this device merely produces an abstract schema too far removed from reality.

For example, in econometrics the price of a commodity is considered to be a function of a whole host of variables, such as the incomes of various social groups, prices of other competing goods, production costs, the nature of the commodity itself (*e.g.* a necessity or a luxury), and the seasonal effects in the case of what are called 'anchovy goods', which are goods with very heavy supply-fluctuations, such as the catch of fish, which in one year may be a hundredfold that of another year. An econometric law which correlates one magnitude with another single variable under the *ceteris paribus* condition (other factors remaining the same) can never be actually tested, as the 'other factors' do not, in fact, remain the same. Nor would it be of much use even if it were 'true' in some imaginary market where other factors were assumed to remain the same. But long before econometricians felt the need for a calculus of multi-variate functions, that is functions depending on more than one independent variable, mathematicians had begun to create one in an attempt to study fluid motions.

Just as Newton was led to the notion of differential coefficients in his studies of the motion of projectiles and heavenly bodies, so also Euler was led to the notion of *partial* differential coefficients in applying Newton's equations of ordinary dynamics to fluid motion. In ordinary dynamics the projectile (or heavenly body) is considered to be a particle whose velocity is a function of the single independent variable time. But a fluid cannot be treated as a mere particle. It has bulk which cannot be disregarded even in an abstract schematic treatment. Moreover, a fluid in general not only varies in velocity from point to point in space at any one instant, but also from moment to moment of time at any one point.

A fluid motion, say on the Earth, has therefore two aspects: a geographic aspect and an historic aspect. When we consider the former, we fix our attention on a particular instant t of time and wish to study the velocity of a fluid particle as a function of its geographic position. In the latter we rivet our attention on a specified particle of the fluid and study its velocity as a function of time. A combination of both the aspects, that is, consideration of the velocity of a fluid particle as a function of both space and time simultaneously, will give an actual picture of the fluid flow as a whole. In this way we are led to the notion of a multivariate function; that is, a function depending on more than one variable. Such, for instance, is the case with the velocity of a fluid particle, as it is a function not only of time but also of the position co-ordinates of the particle in question.

To illustrate the idea of the partial differential coefficient to which Euler was led by a study of fluid motions, let us consider a function y which depends on two independent variables x_1, x_2, simultaneously:

$$y = f(x_1, x_2).$$

Consider first x_1. Since y is a function of x_1, x_2 only, that is, it depends on only x_1, x_2, any change (small or otherwise) in its value can arise only on account of a change in x_1 or x_2 or both. In science, whenever we have to investigate a phenomenon which depends on more than one cause (*e.g.* agricultural yield, which among other things is a function of the qualities of both manure and seed), we disentangle the complex field of influence of the two factors by studying the influence of each in isolation from the other. Once we have studied the effect of the causes singly, we can give due weight to each when the isolates are put back into their natural interrelations.

This method is appropriate in the case under consideration. To begin with, we make a small change dx_1 in the value of x_1 only, making no change in that of x_2. What change would it cause in the value of $f(x_1)$? As we have seen, the derivative df/dx_1, is the rate of change of $f(x_1)$ or y per unit change of x_1. That is, for every unit change of x_1, the corresponding change in f or y is df/dx_1. Hence for a small change dx_1 in the value of x_1 the corresponding change in f or y would be the product:

$$\left(\frac{df}{dx_1}\right)dx_1.$$

Note carefully that f is a function of x_1 as well as x_2; but as we have assumed no change in the value of x_2 we must express this fact somehow in our notation. This is done by writing $\partial f/\partial x_1$ instead of df/dx_1 so as to remind us that the other variable, x_2, is not to be changed and is therefore to be treated *as if* it were a constant. The quantity $\partial f/\partial x_1$ is known as the partial derivative of f with respect to x_1. Hence the change in f or y corresponding to a small change in x_1, x_2 remaining unchanged, is $\left(\frac{\partial f}{\partial x_1}\right) dx_1$.

We now study the change in f (or y) corresponding to a small change in the value of x_2 while no change is made in the value of x_1. A similar argument shows that it is:

$$\frac{\partial f}{\partial x_2}dx_2.$$

What then is the change in f (or y) if both x_1, x_2 are changed simultaneously? We may assume that it is the sum of the changes induced by the small changes in x_1 and x_2 acting singly. It may not be an accurate assumption to make when the changes in x_1 and x_2 are combined to produce a new complex, but it is quite good, for all practical purposes, *if* the changes in x_1, x_2 are infinitesimally small. We therefore conclude that a small change in the value of y, induced by the joint operation of small changes x_1 and x_2 in the values of x_1 and x_2, is the sum of the changes induced by them singly. In other words,

$$dy = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2.$$

In obtaining his equations of motion of fluid flow Euler merely made use of an extension of this theorem, *viz.* a small change in any function $f(x_1, x_2, x_3, t)$ of four independent variables x_1, x_2, x_3, and time t is given by the sum:

$$df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \frac{\partial f}{\partial x_3}dx_3 + \frac{\partial f}{\partial t}dt.$$

With the help of this theorem and the application to a small volume of fluid element the second law of Newton—*viz.*, that the product of the mass and rate of change of velocity in any direction is equal to the resultant of forces (including fluid pressures) acting thereon in that direction, Euler obtained equations of motion connecting the partial differential coefficients (or rates of change) of velocities with respect to the three spatial co-ordinates and the time. In addition, he obtained another partial differential equation—the equation of continuity—from the consideration that the mass of any moving fluid element under consideration remains constant.

Now, since velocity at any point has three components, say (u, v, w), along the three co-ordinate axes, it is obvious that the above-mentioned equations would involve partial differential coefficients of three separate functions (u, v, w) with respect to x, y, z and t. It is possible in many cases to merge the search for the three separate functions, u, v, w, satisfying these equations into the search for a single function called the velocity-potential function, Φ, of four independent variables x, y, z, t satisfying the same equations. But the equations by themselves do not specify the velocity-potential function uniquely. They are satisfied by a far more general class of functions of which the potential function Φ is just one. To narrow down further the margin of indeterminateness we have to make use of boundary conditions which the velocity potential must satisfy. What this means may be explained by an example.

Suppose we consider the flow of water in a canal between two parallel banks. In this case the velocity potential Φ must not only satisfy the Eulerian equations of motion and continuity referred to above but also the further condition that at the boundary of the fluid (that is, at the bank) the velocity component perpendicular to the bank is zero. In other words, Φ must satisfy the boundary condition that the fluid flow at the bank is parallel to it and that no particle flows past it, which could happen only if the banks were breached. Thus, if the Eulerian equations of motion and continuity are the slippers that isolate our Φ-Cinderella from the undesirable members of its family, the boundary condition is the censor that forbids the banns should a relation attempt to masquerade as Φ-Cinderella by wearing the slippers. Unfortunately it has not been possible to discover a Φ-Cinderella who could wear the slippers and satisfy the censor in the general case of fluid motions. And yet Euler had made the slippers rather loose, since he disregarded an important property of real fluids, *viz.* their viscosity.* Seventy years later Navier and Stokes made them much tighter by adding what may be called the *viscous terms* to the three Eulerian equations of fluid motion. That made matters more difficult.

Largely owing to the great inherent difficulties of the subject, hydrodynamical theory was obliged to make a number of simplifying assumptions and thus became more a study of 'perfect' or 'ideal' fluids than that of actual fluids in the real world. This was natural. But unfortunately the next step—that of proceeding to a more realistic state of affairs by embroidering variation from the ideal on the theory—was long delayed. As a result the theoretical development of fluid mechanics did not lead to such perfect harmony between theory and observation as in other branches of mathematical physics such as optics, electricity, magnetism, thermodynamics, *etc.*

For instance, soon after the formulation of the Eulerian equations of fluid motion a paradox, known as D'Alembert's paradox, emerged. The paradox arose because hydrodynamical theory seemed to prove that any body completely immersed in a uniform, steady stream of fluid would experience no resistance whatever—a result quite contrary to experience. D'Alembert's paradox was not an isolated case where plausible hydrodynamical argument led to a conclusion contradicted by physical observation. Before long it appeared that classical hydrodynamical theory was replete with paradoxes which it was unable to rationalise. Such, for instance, were Kopal's paradoxes, paradoxes of airfoil theory, the reversibility paradox, the rising bubble paradox, the paradox of turbulence in pipes, Stokes' paradox, the Eiffel paradox, the Earnshaw paradox, the DuBuat paradox, *etc.*

The emergence of such a multitude of paradoxes clearly showed that hydrodynamical theory did not conform to experimental reality. One

* We commonly call liquids like tar or treacle viscous as they exhibit a tendency to resist change of shape.

consequence of this failure was particularly unfortunate. It made hydrodynamics increasingly abstract, academic and removed from actuality. The engineers, who felt that the mathematicians had left them in the lurch by producing hydrodynamical results largely at variance with reality, began to create a new science of their own. This science of hydraulics was designed to give approximate solutions to real problems. But, lacking a sound theoretical foundation, hydraulics rapidly degenerated into a morass of empirical and semi-empirical formulae. So, by the close of the nineteenth century, both the mathematicians engrossed in 'pure' theory devoid of fresh physical inspiration, and the engineers absorbed in accumulating experimental data without adequate rationalisation by deductive theory, seemed each to have reached blind alleys of their own.

With further progress thus blocked, it was now time to think of a way out. Naturally the very first question that arose was to consider whether these paradoxes were due to the neglect of viscosity or to some more serious flaw deep in the fundamental assumptions of traditional hydrodynamics. We shall examine this basic question of the foundations of hydrodynamics a little more fully. To begin with, let us take viscosity.

It will be recalled that hydrodynamical theory considered *ideal* fluids of zero viscosity rather than *real* fluids which show more or less tendency to resist change of shape. But this is only part of the story. The complete story is that in deriving the equations of fluid motion we must make one of two approximations. Either we neglect viscosity, which leads to Euler's equations for a so-called non-viscous fluid, or we neglect compressibility and assume that fluid density remains constant to obtain Navier-Stokes' equations for an incompressible fluid. Can we attribute the paradoxes of hydrodynamics *solely* to one of these two approximations? The answer is no, because it can be shown that many of the paradoxes are not due to one or other of these two approximations. But even if it were shown that they were, there seems to be no way of avoiding them, for even with these simplifying approximations the mathematical problem in most cases is quite involved if not intractable. Besides, these approximations have to be made because we do not actually know how viscosity acts in a fluid under rapid compression. Without them hydrodynamical theory, therefore, would have to stop almost at the very threshold. How are we then to rationalise hydrodynamical paradoxes?

Birkhoff has recently shown that there can be no simple answer to this difficult question. What is required is a profound analysis of the entire body of mathematical, logical and physical assumptions (tacit or otherwise) of hydrodynamical theory *in the light of experimental data*, for, as Birkhoff has demonstrated, hydrodynamical paradoxes are *not solely* due to the single 'unjustified' neglect of viscosity. They are equally due to faulty physical and logical assumptions underlying hydrodynamical reasoning. Take, for instance, the rising bubble paradox. If we consider a small air bubble rising in a large mass of water under its own buoyancy, conditions of symmetry* require that it should rise vertically. And yet in most cases it ascends in a vertical spiral instead. This paradox arises because of the assumption that 'symmetric causes produce symmetric effects'. But as Birkhoff rightly suggests, the symmetric effects that symmetric causes produce need not necessarily be stable. If they happen to be unstable, a slight deviation would tend to multiply and the symmetry in effects would be too short-lived to be noticeable: the observed effects therefore would be non-symmetric. In other words, while exact symmetric causes would, no doubt, produce exact symmetric effects, nearly symmetric causes need not produce nearly symmetric effects *if they happen to be unstable*. Hence before we can legitimately use the arguments of symmetry, we must first show that symmetric effects deduced are stable. But a demonstration of the stability of a mathematical problem is far more intricate than the deduction itself. It therefore happens that the deduction is often made before its stability can be proved. This leaves only one alternative, *viz.*, to make use of the symmetric argument but to test the stability of deduction in practice—that is, by experiment. This is one illustration of the way in which deductive theory must be interwoven with experimental practice if hydrodynamics is to be freed from paradox.

Another cause of paradox is the assumption that small causes produce small effects. Since every physical experiment is actually affected by innumerable minute causes, we should be quite unable to predict the result of a single such experiment if we did not continually make this assumption; and yet it is not universally true. An obscure fanatic's bullet—as at Sarajevo—may precipitate a global war, a slight fault in the earth's crust may cause a devastating earthquake or a deep-sea explosion, and a single mutation in a gene may alter the entire genetic mould of an individual or even a race. In the limited field of fluid mechanics there are cases where arbitrarily small causes do produce *significant* effects which cannot be neglected. For instance, a small change in viscosity (though not a small change in compressibility) may drastically affect fluid flow. This is because in Navier-Stokes' equations of fluid motion viscosity is the coefficient of the highest order derivative appearing in the equations. Now in systems of differential equations, the presence of arbitrarily small terms of higher order can entirely change the behaviour of the solutions. It is *not* always the case that as the coefficient of a term in an equation tends to zero, its solution tends to the solution obtained by deleting that coefficient. It may also

* Under surface tension the air bubble will be spherical and there is no cause which does not operate symmetrically through the centre of the bubble.

happen that as this coefficient tends to zero the solution suffers an abrupt change of nature at some stage. Thus, for example, in the equations of motion of a sphere through a fluid, a minute change in the value of the viscosity coefficient from a small value of 10^{-5} to $0{\cdot}5 \times 10^{-5}$ causes a sudden and radical change in the nature of the whole solution, resulting in Eiffel's paradox.

We must therefore learn to discriminate between 'right' and 'wrong' approximations. The only way we can do so is boldly to use them in our deductive theory but to *test* the conclusions so derived by experiment. If any of its conclusions is contradicted by physical observation, we should have to examine all the approximations made to discover those at fault. This is the only way we can proceed and even then we may not always be able to overcome the trouble. For it is quite possible for a paradox to arise because *too many* approximations have been assumed* and it is almost impossible to say which of them was 'wrong'.

These considerations show that hydrodynamical paradoxes are due to over-free use of approximations, non-rigorous symmetry considerations and physical over-simplifications. But this does not mean that we can rescue hydrodynamical theory by supplanting mathematical deduction by the more 'physical' reasoning so popular with practical engineers—quite the contrary. For recent developments in fluid mechanics have shown that mathematics is not just a useful device for presenting results whose broad outlines were suggested by physical intuition, as Archimedes and Newton used geometry to present results derived by 'analytical' or 'fluxion' methods. In many cases mathematical deduction gives correct results verified by later experiments which physical intuition is not only unable to derive but would, in fact, straightway reject as grossly absurd. The moral of all this is that what we need to build (that is, a paradox-free fluid mechanics) is a happy blend of the practice of the hydraulic engineer with the deductive theory of the mathematician—a complete interpenetration of deductive theory and practical experience at all levels. It is fortunate that during the past fifty years, mainly as a result of the impetus provided by the needs of aviation, this gap between hydrodynamical theory and experiment has been progressively bridged by just such a happy blend.

* * * *

Lewis Mumford has divided the history of the Western machine civilisation during the past millennium into three successive but over-lapping and interpenetrating phases. During the first phase—the eotechnic phase—trade, which at the beginning was no more than an irregular trickle, grew to such an extent that it transformed the whole life of Western Europe. It

* This is the case with the Earnshaw paradox.

is true that the development of trade led to a steady growth of manufacture as well, but throughout this period (which lasted till about the middle of the eighteenth century) trade on the whole dominated manufacture. Thus it was that the minds of men were occupied more with problems connected with trade, such as the evolution of safe and reliable methods of navigation, than with those of manufacture. Consequently, while the two ancient sources of power, wind and water, were developed at a steadily accelerating pace to increase manufacture, the attention of most leading scientists, particularly during the last three centuries of this phase, was directed towards the solution of navigational problems. The chief and most difficult of these was that of finding the longitude of a ship at sea. It was imperative that a solution should be found as the inability to determine longitudes led to very heavy shipping losses. Newton had tackled it, although without providing a satisfactorily practical answer. In fact, as Hessen has shown, Newton's masterpiece, the *Principia*, was in part an endeavour to deal with the problems of gravity, planetary motions and the shape and size of the earth, in order to meet the demands for better navigation. It was shown that the most promising method of determining longitude from observation of heavenly bodies was provided by the moon. The theory of lunar motion, therefore, began to absorb the attention of an increasing number of distinguished mathematicians of England, France, Germany and America.

Although more arithmetic and algebra were devoted to Lunar Theory than to any other question of astronomy or mathematical physics, a solution was not found till the middle of the eighteenth century, when successful chronometers, that could keep time on a ship in spite of pitching and rolling in rough weather, were constructed. Once the problem of longitude was solved it led to a further growth of trade, which in turn induced a corresponding increase in manufacture. A stage was now reached when the old sources of power, namely wind and water, proved too 'weak, fickle, and irregular' to meet the needs of a trade that had burst all previous bounds. Men began to look for new sources of power rather than new trade routes.

This change marks the beginning of Mumford's second phase, the palaeotechnic phase, which ushered in the era of the 'dark Satanic mills'. As manufacture began to dominate trade, the problem of discovering new prime movers became the dominant social problem of the time. It was eventually solved by the invention of the steam engine. The discovery of the power of steam—the chief palaeotechnic source of power—was not the work of 'pure' scientists; it was made possible by the combined efforts of a long succession of technicians, craftsmen and engineers from Porta, Rivault, Caus, Branca, Savery and Newcomen to Watt and Boulton.

Although the power of steam to do useful work had been known since the time of Hero of Alexandria (A.D. 50), the social impetus to make it the chief prime mover was lacking before the eighteenth century. Further, a successful steam engine could not have been invented even then had it not been for the introduction by craftsmen of more precise methods of measurement in engineering design. Thus, the success of the first two engines that Watt erected at Bloomfield colliery in Staffordshire, and at John Wilkinson's new foundry at Broseley, depended in a great measure on the accurate cylinders made by Wilkinson's new machine tool with a limit of error not exceeding 'the thickness of a thin six pence' in a diameter of seventy-two inches. The importance of the introduction of new precision tools, producing parts with increasingly narrower 'tolerances', in revolutionising production has never been fully recognised. The transformation of the steam engine from the wasteful burner of fuel that it was at the time of Newcomen into the economical source of power that it became in the hands of Watt and his successors seventy years later, was achieved as much by the introduction of precision methods in technology as by Black's discovery of the latent heat of steam.

A natural consequence of the introduction of higher standards of refinement in industry and technology was that mathematical language itself became increasingly exact, subtle, fine, intricate and complex. The greatest change in this direction occurred in the language of the infinitesimal calculus, as the reasoning on which its technique had been based was shockingly illogical.

The methods of calculus were accepted, not because their reasoning was logically impeccable, but because they 'worked'—that is, led to useful results. For instance, in calculating the 'fluxion', or—as we now say—the differential coefficient of x^2, the founders of the calculus would substitute $x + 0$ for x in a term like x^2, expand the resultant expression $(x + 0)^2$ as if the zero within the bracket was a non-vanishing quantity, and then let 0 disappear in the final step. In other words, they believed that there existed quantities known as the 'infinitesimals', or 'fluxions', which could be treated as zero or non-zero according to the convenience of the mathematician.

This glaring illogicality of the calculus did not escape unnoticed. It was exposed with masterly skill by a non-mathematician, Bishop Berkeley, the famous idealist philosopher. While the mathematicians, with unerring instinct, ignored the attack and went on piling formulae upon formulae like Ossa upon Pelion, these methods brought the mathematicians into bad repute. Swift's caricature of the mathematicians of Laputa and his denunciation of them as 'very bad reasoners' was possibly inspired by Berkeley's withering critique of the 'fluxions'. In his *L'Homme Aux*

Quarante Ecus, Voltaire, too, had a dig at the analysts when he remarked that a 'geometer shows you that between a circle and a tangent you can pass an infinity of curved lines but only one straight line, while your eyes and reason tell you otherwise.'

It is true that some gifted mathematicians of the eighteenth century had realised that although the calculus towered over mathematics like a colossus, its feet were made of clay. Thus, D'Alembert neatly epitomised the whole situation when he remarked that 'the theory of limit is the true metaphysics of the differential calculus'. Lagrange, another eminent mathematician of the same century, tried, though unsuccessfully, to cut the Gordian knot by jettisoning the infinitesimals, 'fluxions' and 'limits' as so much useless lumber, and by representing all functions as sums of a power series—that is, a non-terminating series like $a_0 + a_1x + a_2x^2 + \ldots$. No doubt, in this way he escaped the mysterious infinitesimals which were both zero and not zero at the same time, but he thereby ran into another difficulty, no less serious—that of summing an infinite series.

The way out of the difficulty was first pointed out by a French mathematician, Cauchy, who, during the second decade of the nineteenth century, set out to purge the calculus of all its illogicalities, loose methods of reasoning, and premises of doubtful validity. He thus virtually introduced a 'New Look' that has come to stay in mathematical reasoning. New Look was suspicious of all arguments based on vague analogies and geometric intuition. So it began to examine more precisely all those vague notions and concepts which had hitherto been taken for granted. Take, for instance, the idea of mathematical limit which we explained earlier. The founders of the calculus thought they knew what they meant by a limit. In ordinary language, we use it to mean a sort of a terminal point or a bound that may not or cannot be passed. Thus, the Phoenician navigators considered the Pillars of Hercules as the *limit* of navigation, because in those days few ships that sailed beyond them ever returned. When mathematicians began to define speed *at* an instant as the limit of the average speed when the time interval over which it is averaged tends to zero, they pictured it in much the same manner as the Phoenicians thought of the Pillars—as a sort of barrier past which the average cannot go. But this explanation of the limit notion, though it may be a good enough start, is not sufficient. For, while the mathematical meaning of limit is vaguely suggested by its linguistic usage, it is not precisely defined thereby. Cauchy gave the first genuinely mathematical definition of limit, and it has never required modification. The need for it had been realised earlier by many, but it came into being nearly 150 years after mathematicians had been manipulating with limits. So if you do not get it right first reading, do not be disappointed. It took the mathematicians themselves the best part of a century to frame it.

To grasp the idea underlying Cauchy's definition of a limit, let us consider once again the distance function $y = t^2$ by which we introduced the limit notion, but to give greater generality to our results we shall make our symbols y and t meaning-free. That is, we shall no longer think of y as distance and t as time but let them denote any two variables whatever related by the same functional formula. It would help in this abstraction if we changed the notation a little and worked with the formula $y = x^2$ instead of $y = t^2$. Earlier we defined dy/dx at $x = 2$ as the *limit* of

$$\frac{(2 + h)^2 - (2)^2}{(2 + h) - 2} \quad . \quad . \quad . \quad . \quad (1)$$

when h tends to zero. Here the expression (1) is really a function of h as it depends on the value of h taken. For any given value of h (except $h = 0$*), we can calculate the corresponding value of (1) by substituting it for h in the expression (1). In fact, if we simplify this expression, we find that it is

$$\frac{4 + 4h + h^2 - 4}{h} = \frac{(4 + h)h}{h} = 4 + h$$

for *any* value of h other than zero.

This shows that if we assume h successively to be ·1, ·01, ·001, ·0001, . . ., the corresponding values of (1) are 4·1, 4·01, 4·001, 4·0001, . . . respectively.

We thus conclude that the limit of (1) is 4 as h is indefinitely decreased. This may seem to suggest that the limiting number 4 is a sort of barrier point beyond which the value of (1) cannot pass. Actually this is not always true. Thus if we assign h a succession of negative and positive values ·1, —·1, ·01, —·01, ·001, —·001, ·0001, —·0001, . . . we obtain for (1) the values 4·1, 3·9, 4·01, 3·99, 4·001, 3·999, 4·0001, 3·9999, ...

Here even though the value of (1) continually crosses the barrier number 4 backward and forward, nevertheless it keeps on coming closer and closer to 4. We are, therefore, still justified in calling 4 as the limit of (1). The essential idea of the limit is not that it is a sort of impassable terminal but that it is a point of continually closer and closer approach. For this purpose it is, of course, necessary that we should be able to make the value of (1) approach as near its limiting value 4 as we like by making h sufficiently small. But what is even more important is that once we have *so*

* For $h = 0$, the formula gives the meaningless expression 0/0. Division by zero is a taboo in mathematics as it leads to *no* result. For instance, if you had hundred rupees you could issue ten cheques of Rs. 10 each in all. But if you issued cheques of zero value, you could issue *any* number of them for all the difference it would make to your bank balance.

brought the value of (1) *within any arbitrarily chosen small range of closeness to its limiting value, it should continue to remain as close, if not closer, for all other numerically smaller values of h.* Thus, suppose we wish the value (1) to differ from its limiting value 4 by less than an arbitrarily small number, say, ε. In order that the difference

$$(4 + h) - 4$$

may remain numerically less than ε, all that we need do is to make h less than ε. So long as we keep h smaller than ε the difference between 4 and the actual value of the expression (1) will never exceed ε. In other words, the value of (1) always lies between $4 - \varepsilon$ and $4 + \varepsilon$ for *all* values of h less than ε.

In general, we say that a function $f(x)$ tends to a limit l as x tends to zero, if we can *keep* $f(x)$ as close to l as we like by keeping x sufficiently close to zero. This means that given any arbitrarily small number ε another small number δ depending on ε alone can be found such that the difference $l - f(x)$ *always* remains numerically below ε as long as x stays below δ. In other words, if you wish to imprison *all* the values of $f(x)$ within the two corners of a number mesh of any given degree of narrowness—no matter how small—round the number l, you have to devise *another* number mesh round zero within which the value of x must remain confined. We usually denote the narrowness of the number mesh round l by ε, where ε is *any* arbitrarily small number we care to nominate.

In the same way we could examine if $f(x)$ tends to a definite limit, say, L when x tends to any other value, say, $x = a$. The examination may be done in steps as follows. In the first place, we nominate an arbitrarily small number to specify the narrowness of the number mesh round L within which we wish to trap all values of $f(x)$. Let it be ε. This means that we wish to confine all values of $f(x)$ within the number mesh $L - \varepsilon, L + \varepsilon$.

The second step is to discover what restriction should be imposed on the values of x in order to confine $f(x)$ within $L - \varepsilon, L + \varepsilon$. This means that we are to examine whether we can discover another small number δ depending on ε, such that if we confine x to the number mesh, $(a - \delta, a + \delta)$ round a, the function $f(x)$ remains within $(L - \varepsilon, L + \varepsilon)$. If this condition is satisfied, we say $f(x)$ tends to L as x tends to a.

This is Cauchy's famous 'epsilon-delta' definition of limit; he defined it in purely arithmetical terms and thus freed it from vague associations with its counterpart of everyday speech.

* * * *

The motive force of modern mathematics is abstraction: in fact, abstraction *is* power and the reason is this. Whenever we treat some real problem mathematically, whether in physics, astronomy, biology or the social sciences, there is only one way in which we can proceed. We must first simplify it by having recourse to some sort of an abstract model or replica representing those features of reality considered most essential for the problem in question. Take, for instance, Eddington's famous elephant problem. Here an elephant weighing two tons slides down a grassy hill-side of 60° slope. How long did he take to slide down? If we strip the problem of its 'poetry' (Eddington's word), that is, if we make an abstract mathematical model embodying its essential features, all that we have is a 'particle' sliding down an 'inclined plane'. Here 'particle' is a mathematical abstraction, which just retains that essential property of material bodies we call 'ponderosity' or 'inertia' common to them all. Similarly, 'inclined plane' is a geometrical abstraction of hillsides embodying its essential feature 'steepness'.

Abstract mathematical models of this kind not only simplify the real problem by retaining only the bare essentials without those encumber-ances, the irrelevant details which complicate matters and make it in-tractable, but also apply to a much wider class of problems than the original. This explains the paradoxical statement sometimes made: the more modern mathematics departs from reality (that is, grows abstract) the closer it comes to it. For no matter how abstract it may appear, it is in the ultimate analysis an embodiment of certain essential features abstracted from some sphere of reality.

Now an important group of problems in the most varied fields of science may be reduced to the consideration of systems changing with time. We have already given instances of it—the motion of material bodies such as stones and elephants rolling down grassy hill-sides, and the changing posi-tions of planets in the sky. Other instances of this kind are: in biology, the development of a population or the growth of an organism; in astronomy, the changing luminosity of variable stars; in economics, the varying de-mands and prices of a market; in actuarial science, the changing number of claims for payment on an insurance company; in telephone theory, the changing incidence of telephone calls, etc. In any of these problems we may imagine an abstract mathematical model where 'something' changes with time. In some problems, such as the development of a biological population or the growth of insurance claims, this 'something' changes at discrete moments and in whole numbers. That is, it jumps from one whole number to another.* In others, such as the positions of planets in the sky or the growth of a biological organism, the essential feature of the change is that it takes place incessantly by small degrees.

* For instance, if the population of a species is, say, 10 at any given moment, it cannot possibly attain fractional values like 10·5 or 9·2 at any other time. The birth and death of its individuals can only alter it in whole numbers and at discrete moments.

There are thus two main kinds of mathematical models. In the one designed for handling continuous processes, this 'something', or the dependent variable (y), changes continuously with time. In the other, designed for treating discontinuous processes, it *jumps* discontinuously. We could, in fact, carry the process of abstraction still farther and consider a set of related variables—a dependent variable y depending on an independent variable x which need not necessarily represent time. The application of the fundamental laws of the process usually leads to one or more differential, integral or integro-differential equations. The question then arises whether such equations can be solved. In many cases it can be proved that there is a unique solution, provided it is assumed that y is a continuous function of x. It is, therefore, important to examine under what conditions a function is continuous.

Our naïve intuitive idea of a continuous function is this. It is continuous when its graph can be drawn without ever lifting the pencil from paper. If the graph jumps at any particular point, the function is discontinuous at that point. Consider, for instance, a rocket fired vertically upwards with an initial speed of, say 440 feet per second. Suppose further that the rocket carries an explosive charge, which on explosion during the course of its ascent is enough to impart to it instantaneously a further vertical speed of 88 feet per second. If the charge explodes, say five seconds after the initial start, the speed function (y) will be given by the formulae

$$y = 440 - 32x, \quad . \quad . \quad . \quad . \quad (1)$$

for all x up to and equal to 5,

and,
$$y = 528 - 32x, \quad . \quad . \quad . \quad . \quad (2)$$

for all values exceeding 5 till the vertical ascent of the rocket ceases.

If we draw its graph, it will be represented by the line AB for the values of x lying between 0 and 5. (See Fig. 8.) At $x = 5$ the rocket charge explodes and increases its speed instantaneously by 88 feet per second. The speed graph is now represented by the parallel line CD for values of x exceeding 5. The speed function as a whole is, therefore, represented by the two lines AB and CD. At the point B we have to lift the pencil to draw it. This shows that the speed function is discontinuous at B.

We can make this notion of continuity more rigorous, if abstract, by disregarding the graph of the function and considering only the functional relation between y and x. If we are considering a range of values of x from 0 to 5, formula (1) applies. Hence if x tends to 5 from below, that is, by always remaining less than 5, y tends to $440 - 32(5) = 280$. On the other hand, if x tends to 5 from above, that is, by remaining always more than 5, formula (2) applies and the value of y tends to $528 - 32(5) = 368$.

There is thus a hiatus at $x = 5$. If we approach from below y tends to 280, while if we approach it from above it tends to 368. This is merely a mathematical translation of the physical fact that at the instant $x = 5$ the speed of the rocket jumped by 88 feet per second due to the instantaneous explosion of the charge during its ascent.

Whenever we have a function whose value tends to two different limits as x tends to any given value a from above and below, the function is said

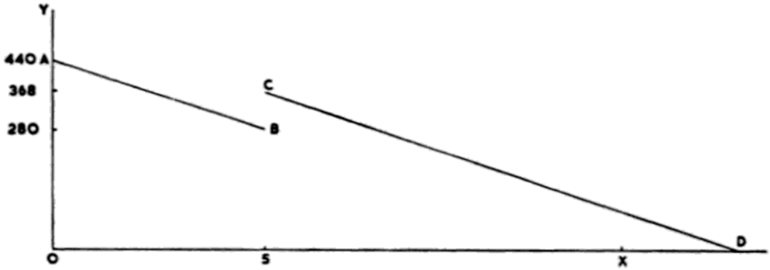

FIG. 8—The speed function of the rocket jumps at B to C and is therefore discontinuous at $x = 5$.

to be discontinuous at $x = a$. We thus have three values of a function at any point $x = a$: its actual value at $x = a$ and the two limits as x approaches a from two sides—above and below. The function is continuous at $x = a$ if and only *if* all three are equal. If *any* two happen to differ, the function is discontinuous at $x = a$.

All ordinary functions defined by one or more formulae are generally continuous except, possibly, for some isolated values. Thus, for instance, the speed function cited above was continuous everywhere except at $x = 5$. Is it possible to have functions defined by regular mathematical formulae, which are discontinuous everywhere throughout the range of the independent variable? The earlier mathematicians would have unhesitatingly replied no. We now know better. Suppose we have a function $y = f(x)$ defined by the following two formulae for all values of x in the range (0, 1):

$$y = f(x) = 0, \text{ whenever } x \text{ is a rational fraction;}$$

and

$$y = f(x) = 1, \text{ whenever } x \text{ is an irrational fraction.}$$

Thus for $x = \frac{1}{2}$, y is zero; but for $x = \sqrt{\frac{1}{2}}$ y is unity.

Such a function is everywhere *discontinuous*. For, take *any* value of x such as $x = \frac{1}{2}$. Within any range round $\frac{1}{2}$ no matter how small, there are any number of rational and irrational values. Thus in the range $(\frac{1}{2} - \frac{1}{8000}, \frac{1}{2} + \frac{1}{8000})$, the value $\frac{1}{2} + \frac{1}{8000}$ is rational and $\frac{1}{2} + \frac{1}{8000\sqrt{2}}$ is irrational.

For the rational values of x in the range, $f(x)$ will be zero while for its irrational values $f(x) = 1$.

The function $f(x)$ therefore continually oscillates between 0 and 1 however narrow the range of values of x round $x = \frac{1}{2}$ we may choose. It cannot, therefore, tend to any limit as x tends to $\frac{1}{2}$. But there is nothing in the above argument special about $x = \frac{1}{2}$. What has been shown about the limiting behaviour of $f(x)$ when x approaches $\frac{1}{2}$ holds equally for *any* other value of x in the interval (0, 1) for which we have defined the function $f(x)$. The function $f(x)$ is therefore discontinuous *everywhere*. It gives us an absolute discontinuity.

If we have functions which are discontinuous somewhere as well as everywhere, what about their differential coefficients at these points of dis-

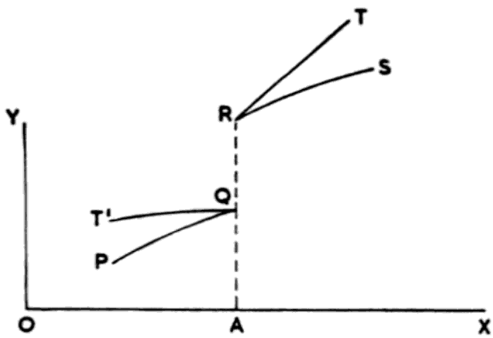

FIG. 9—Here $f(x)$ is discontinuous. As x approaches A the graph line suddenly jumps from Q to R. At this point no single line which can be regarded as a tangent to the graph line as a whole can be drawn. The derivative of $f(x)$, therefore, does not exist at A.

continuity? Clearly a discontinuous function can have no differential coefficient at a point of discontinuity. For, if we draw the graph of the function (see Fig. 9), there will be a sudden jump at a point of discontinuity, so that it is impossible to draw a tangent to it at that point. But since the slope or gradient of the tangent is the measure of the differential coefficient, the latter cannot exist if the tangent cannot be drawn. If the tangent cannot be drawn at a point of discontinuity, can it be drawn at all points of continuity? Not necessarily. Examine, for instance, a continuous curve like the one drawn in Fig. 10. It is continuous everywhere even at the kinky point Q. But no tangent can be drawn to the graph as a whole at Q. For the portion of the graph to the right of Q, the tangent at Q is QT, for that on the left, QT'. Since there is no unique tangent to the graph at Q, the differential coefficient cannot exist although the graph is continuous.

It might appear that Q is an exception and that in general a continuous function is bound to have a tangent everywhere except possibly at a few isolated points. This, however, is one instance where our geometric intuition leads us astray. For we can construct functions which, though continuous for *every* value of the independent variable, have no derivative for

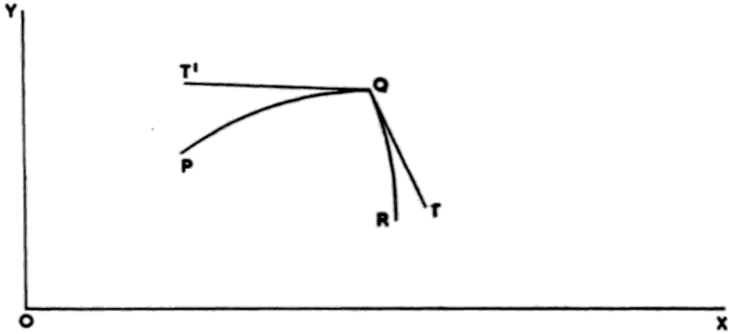

FIG. 10—The graph line suddenly turns turtle at Q Although the graph line is clearly continuous at Q, no single tangent can be drawn to it there. Hence the derivative of $f(x)$ does not exist at Q.

any value whatever. If we draw the graph of any such function, it will be a continuous line at *no* point of which we can draw a tangent. Every point on the graph will thus be a kink like the point Q in Fig. 9. That is why it is impossible to draw it on paper.

* * * *

A similar attempt was made to rigourise and refine the theory of integration. The refinement, no doubt, made it more abstract, complex and subtle, but it was worthwhile. It paved the way for a much wider generalisation of the theory of integration, which has been extensively applied. We shall deal with this generalisation and its applications in Chapter 6.

To understand the refinement, let us recall the essential steps in the calculation of the speed integral $y = t$ over the time interval (1, 6). First, we divided the interval (1, 6) into a large number of equal sub-intervals. Second, we multiplied the duration or length of each sub-interval by the lower and upper limits of the speed function *in* that sub-interval. We thus obtained the lower and upper limits within which lay the actual distance travelled during that sub-interval. Third, we added the products corresponding to all these sub-intervals and formed two sums, the lower sum s and the upper sum S, between which lay the total actual distance travelled. Fourth, we computed the limits of s and S as the number of sub-intervals was indefinitely increased. This common limit—the speed integral over time—was then the actual distance travelled.

GREAT IDEAS OF MODERN MATHEMATICS

JAGJIT SINGH

This book will provide the educated layman with an understanding of the main mathematical ideas used in the sciences of modern civilization. But these are not the elementary ideas which one usually finds in similar books. Though Singh takes up the number system and the calculus, he devotes most of his time to such material as differential equations, matrices, group theory, transformations, logic, statistics, and some of the current problems of mathematical foundations.

If the reader will bring to the book a fair knowledge of high school mathematics, he will find himself propelled forward in the subject by a remarkable teacher. He will understand difficult concepts better than many who use them as tools but comprehend little of their basic structure and design. The author does not hold that mathematics is a game played with meaningless symbols, but constantly shows, by references and illustrations, how mathematical ideas are derived from and related to the physical sciences and man's thinking methods.

Dr. Singh writes smoothly, and in making his points is aided by a wide knowledge of non-mathematical fields. His exposition of a two-valued logic, fundamental to electronic computers, is a splendid example of his lucidity, as are his treatments of imaginary numbers, vectors, probabilities, and statistics. In covering the major mathematical ideas of the past two hundred years, the author enables the layman to begin to understand the workings of modern physics, astronomy, genetics, psychology, and evolution.

An original Dover publication. Author's foreword. Appendixes. Name and subject indexes. 65 illustrations. x + 312pp. 5⅜ x 8.

Paperbound

Cover designed by Theodore Menten

ISBN 0-486-20587-8

90000

9 780486 205878