HOME UNIVERSITY LIBRARY
OF MODERN KNOWLEDGE

AN INTRODUCTION TO MATHEMATICS

By A. N. WHITEHEAD, Sc.D., F.R.S.

London
WILLIAMS & NORGATE

HENRY HOLT & Co., New York
Canada: WM. BRIGGS, Toronto
India: R. & T. WASHBOURNE, Ltd.

HOME UNIVERSITY LIBRARY

OF

MODERN KNOWLEDGE

—

NEW YORK

HENRY HOLT AND COMPANY

AN

INTRODUCTION

TO

MATHEMATICS

BY

A. N. WHITEHEAD,

Sc.D., F.R.S.,

AUTHOR OF "UNIVERSAL ALGEBRA," JOINT
AUTHOR OF "PRINCIPIA MATHEMATICA"

NEW AND REVISED EDITION

LONDON
WILLIAMS AND NORGATE

PRINTED BY
HAZELL, WATSON AND VINEY, LD.,
LONDON AND AYLESBURY.

238

CONTENTS

AN INTRODUCTION TO MATHEMATICS

CHAPTER I

THE ABSTRACT NATURE OF MATHEMATICS

The study of mathematics is apt to commence in disappointment. The important applications of the science, the theoretical interest of its ideas, and the logical rigour of its methods, all generate the expectation of a speedy introduction to processes of interest. We are told that by its aid the stars are weighed and the billions of molecules in a drop of water are counted. Yet, like the ghost of Hamlet's father, this great science eludes the efforts of our mental weapons to grasp it—"'Tis here, 'tis there, 'tis gone"—and what we do see does not suggest the same excuse for illusiveness as sufficed for the ghost, that it is too noble for our gross methods. "A show of violence," if ever excusable, may surely be "offered" to the trivial results which occupy the pages of some elementary mathematical treatises.

The reason for this failure of the science to live up to its reputation is that its fundamental ideas are not explained to the student disentangled from the technical procedure which has been invented to facilitate their exact presentation in particular instances. Accordingly, the unfortunate learner finds himself struggling to acquire a knowledge of a mass of details which are not illuminated by any general conception. Without a doubt, technical facility is a first requisite for valuable mental activity: we shall fail to appreciate the rhythm of Milton, or the passion of Shelley, so long as we find it necessary to spell the words and are not quite certain of the forms of the individual letters. In this sense there is no royal road to learning. But it is equally an error to confine attention to technical processes, excluding consideration of general ideas. Here lies the road to pedantry.

The object of the following Chapters is not to teach mathematics, but to enable students from the very beginning of their course to know what the science is about, and why it is necessarily the foundation of exact thought as applied to natural phenomena. All allusion in what follows to detailed deductions in any part of the science will be inserted merely for the purpose of example, and care will be taken to make the general argument comprehensible, even if here and there some technical process or symbol which the reader does not understand is cited for the purpose of illustration.

The first acquaintance which most people have with mathematics is through arithmetic. That two and two make four is usually taken as the type of a simple mathematical proposition which everyone will have heard of. Arithmetic, therefore, will be a good subject to consider in order to discover, if possible, the most obvious characteristic of the science. Now, the first noticeable fact about arithmetic is that it applies to everything, to tastes and to sounds, to apples and to angels, to the ideas of the mind and to the bones of the body. The nature of the things is perfectly indifferent, of all things it is true that two and two make four. Thus we write down as the leading characteristic of mathematics that it deals with properties and ideas which are applicable to things just because they are things, and apart from any particular feelings, or emotions, or sensations, in any way connected with them. This is what is meant by calling mathematics an abstract science.

The result which we have reached deserves attention. It is natural to think that an abstract science cannot be of much importance in the affairs of human life, because it has omitted from its consideration everything of real interest. It will be remembered that Swift, in his description of Gulliver's voyage to Laputa, is of two minds on this point. He describes the mathematicians of that country as silly and useless dreamers, whose attention has to be awakened by flappers. Also, the mathematical tailor measures his height by a quadrant, and deduces his other dimensions by a rule and compasses, producing a suit of very ill-fitting clothes. On the other hand, the mathematicians of Laputa, by their marvellous invention of the magnetic island floating in the air, ruled the country and maintained their ascendency over their subjects. Swift, indeed, lived at a time peculiarly unsuited for gibes at contemporary mathematicians. Newton's *Principia* had just been written, one of the great forces which have transformed the modern world. Swift might just as well have laughed at an earthquake.

But a mere list of the achievements of mathematics is an unsatisfactory way of arriving at an idea of its importance. It is worth while to spend a little thought in getting at the root reason why mathematics, because of its very abstractness, must always remain one of the most important topics for thought. Let us try to make clear to ourselves why explanations of the order of events necessarily tend to become mathematical.

Consider how all events are interconnected. When we see the lightning, we listen for the thunder; when we hear the wind, we look for the waves on the sea; in the chill autumn, the leaves fall. Everywhere order reigns, so that when some circumstances have been noted we can foresee that others will also be present. The progress of science consists in observing these interconnections and in showing with a patient ingenuity that the events of this evershifting world are but examples of a few general connections or relations called laws. To see what is general in what is particular and what is permanent in what is transitory is the aim of scientific thought. In the eye of science, the fall of an apple, the motion of a planet round a sun, and the clinging of the atmosphere to the earth are all seen as examples of the law of gravity. This possibility of disentangling the most complex evanescent circumstances into various examples of permanent laws is the controlling idea of modern thought.

Now let us think of the sort of laws which we want in order completely to realize this scientific ideal. Our knowledge of the particular facts of the world around us is gained from our sensations. We see, and hear, and taste, and smell, and feel hot and cold, and push, and rub, and ache, and tingle. These are just our own personal sensations: my toothache cannot be your toothache, and my sight cannot be your sight. But we ascribe the origin of these sensations to relations between the things which form the external world. Thus the dentist extracts not the toothache but the tooth. And not only so, we also endeavour to imagine the world as one connected set of things which underlies all the perceptions of all people. There is not one world of things for my sensations and another for yours, but one world in which we both exist. It is the same tooth both for dentist and patient. Also we hear and we touch the same world as we see.

It is easy, therefore, to understand that we want to describe the connections between these external things in some way which does not depend on any particular sensations, nor even on all the sensations of any particular person. The laws satisfied by the course of events in the world of external things are to be described, if possible, in a neutral universal fashion, the same for blind men as for deaf men, and the same for beings with faculties beyond our ken as for normal human beings.

But when we have put aside our immediate sensations, the most serviceable part—from its clearness, definiteness, and universality—of what is left is composed of our general ideas of the abstract formal properties of things; in fact, the abstract mathematical ideas mentioned above. Thus it comes about that, step by step, and not realizing the full meaning of the process, mankind has been led to search for a mathematical description of the properties of the universe, because in this way only can a general idea of the course of events be formed, freed from reference to particular persons or to particular types of sensation. For example, it might be asked at dinner: "What was it which underlay my sensation of sight, yours of touch, and his of taste and smell?" the answer being "an apple." But in its final analysis, science seeks to describe an apple in terms of the positions and motions of molecules, a description which ignores me and you and him, and also ignores sight and touch and taste and smell. Thus mathematical ideas, because they are abstract, supply just what is wanted for a scientific description of the course of events.

This point has usually been misunderstood, from being thought of in too narrow a way. Pythagoras had a glimpse of it when he proclaimed that number was the source of all things. In modern times the belief that the ultimate explanation of all things was to be found in Newtonian mechanics was an adumbration of the truth that all science as it grows towards perfection becomes mathematical in its ideas.

CHAPTER II

VARIABLES

MATHEMATICS as a science commenced when first someone, probably a Greek, proved propositions about *any* things or about *some* things, without specification of definite particular things. These propositions were first enunciated by the Greeks for geometry; and, accordingly, geometry was the great Greek mathematical science. After the rise of geometry centuries passed away before algebra made a really effective start, despite some faint anticipations by the later Greek mathematicians.

The ideas of *any* and of *some* are introduced into algebra by the use of letters, instead of the definite numbers of arithmetic. Thus, instead of saying that $2+3=3+2$, in algebra we generalize and say that, if x and y stand for *any* two numbers, then $x+y=y+x$. Again, in the place of saying that $3>2$, we generalize and say that if x be *any* number there exists *some* number (or numbers) y such that $y>x$. We may remark in passing that this latter assumption—for when put in its strict ultimate form it is an assumption—is of vital importance, both to philosophy and to mathematics; for by it the notion of infinity is introduced. Perhaps it required the introduction of the arabic numerals, by which the use of letters as standing for definite numbers has been completely discarded in mathematics, in order to suggest to mathematicians the technical convenience of the use of letters for the ideas of *any* number and *some* number. The Romans would have stated the number of the year in which this is written in the form MDCCCCX., whereas we write it 1910, thus leaving the letters for the other usage. But this is merely a speculation. After the rise of algebra the differential calculus was invented by Newton and Leibniz, and then a pause in the progress of the philosophy of mathematical thought occurred so far as these notions are concerned; and it was not till within the last few years that it has been realized how fundamental *any* and *some* are to the very nature of mathematics, with the result of opening out still further subjects for mathematical exploration.

Let us now make some simple algebraic statements, with the object of understanding exactly how these fundamental ideas occur.

(1) For *any* number x, $x+2=2+x$;
(2) For *some* number x, $x+2=3$;
(3) For *some* number x, $x+2>3$.

The first point to notice is the possibilities contained in the meaning of *some*, as here used. Since $x+2=2+x$ for any number x, it is true for *some* number x. Thus, as here used, *any* implies *some* and *some* does not exclude *any*. Again, in the second example, there is, in fact, only one number x, such that $x+2=3$, namely only the number 1. Thus the *some* may be one number only. But in the third, example, any number x which is greater than 1 gives $x+2>3$. Hence there are an infinite number of numbers which answer to the *some* number in this case. Thus *some* may be anything between *any* and *one only*, including both these limiting cases.

It is natural to supersede the statements (2) and (3) by the questions:

(2′) For what number x is $x+2=3$;

(3′) For what numbers x is $x+2>3$.

Considering (2′), $x+2=3$ is an equation, and it is easy to see that its solution is $x=3-2=1$. When we have asked the question implied in the statement of the equation $x+2=3$, x is called the unknown. The object of the solution of the equation is the determination of the unknown. Equations are of great importance in mathematics, and it seems as though (2′) exemplified a much more thoroughgoing and fundamental idea than the original statement (2). This, however, is a complete mistake. The idea of the undetermined "variable" as occurring in the use of "some" or "any" is the really important one in mathematics; that of the "unknown" in an equation, which is to be solved as quickly as possible, is only of subordinate use, though of course it is very important. One of the causes of the apparent triviality of much of elementary algebra is the preoccupation of the text-books with the solution of equations. The same remark applies to the solution of the inequality (3′) as compared to the original statement (3).

But the majority of interesting formulæ, especially when the idea of *some* is present, involve more than one variable. For example, the consideration of the pairs of numbers x and y (fractional or integral) which satisfy $x+y=1$ involves the idea of two correlated variables, x and y. When two variables are present the same two main types of statement occur. For example, (1) for *any* pair of numbers, x and y, $x+y=y+x$, and (2) for *some* pairs of numbers, x and y, $x+y=1$.

The second type of statement invites consideration of the aggregate of pairs of numbers which are bound together by some fixed relation—in the case given, by the relation $x+y=1$. One use of formulæ of the first type, true for *any* pair of numbers, is that by them formulæ of the second type can be thrown into an indefinite number of equivalent forms. For example, the relation $x+y=1$ is equivalent to the relations

$$y+x=1, \quad (x-y)+2y=1, \quad 6x+6y=6,$$

and so on. Thus a skilful mathematician uses that equivalent form of the relation under consideration which is most convenient for his immediate purpose.

It is not in general true that, when a pair of terms satisfy some fixed relation, if one of the terms is given the other is also definitely determined. For example, when x and y satisfy $y^2=x$, if $x=4$, y can be ± 2, thus, for any positive value of x there are alternative values for y. Also in the relation $x+y>1$, when either x or y is given, an indefinite number of values remain open for the other.

Again there is another important point to be noticed. If we restrict ourselves to positive numbers, integral or fractional, in considering the relation $x+y=1$, then, if either x or y be greater than 1, there is no positive number which the other can assume so as to satisfy the relation. Thus the "field" of the relation for x is restricted to numbers less than 1, and similarly for the "field" open to y. Again, consider integral numbers only, positive or negative, and take the relation

$y^2 = x$, satisfied by pairs of such numbers. Then whatever integral value is given to y, x can assume one corresponding integral value. So the "field" for y is unrestricted among these positive or negative integers. But the "field" for x is restricted in two ways. In the first place x must be positive, and in the second place, since y is to be integral, x must be a perfect square. Accordingly, the "field" of x is restricted to the set of integers 1^2, 2^2, 3^2, 4^2, and so on, *i.e.*, to 1, 4, 9, 16, and so on.

The study of the general properties of a relation between pairs of numbers is much facilitated by the use of a diagram constructed as follows:

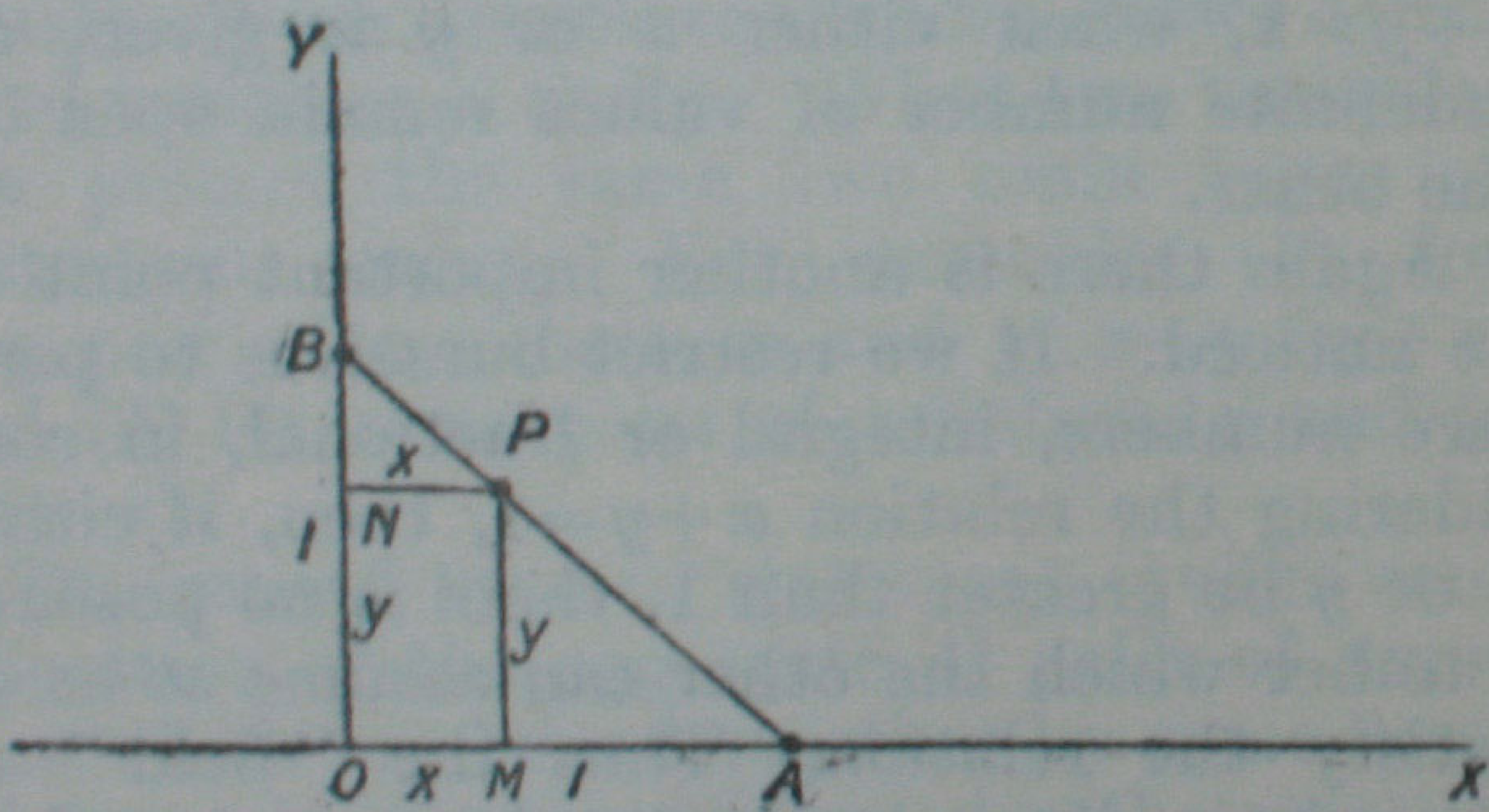


Fig. 1.

Draw two lines OX and OY at right angles; let any number x be represented by x units (in any scale) of length along OX, any number y by y units (in any scale) of length along OY. Thus if OM, along OX, be x units in length, and ON, along OY, be y units in length, by completing the parallelogram $OMPN$ we find a point P which corresponds to the pair of numbers x and y. To each point there corresponds one pair of numbers, and to each pair of numbers there corresponds one point. The pair of numbers are called the coordinates of the point. Then the points whose coordinates satisfy some fixed relation can be indicated in a convenient way, by drawing a line, if they all lie on a line, or by shading an area if they are all points in the area. If the relation can be represented by an equation such as $x+y=1$, or $y^2=x$, then the points lie on a line, which is straight in the former case and curved in the latter. For example, considering only positive numbers, the points whose coordinates satisfy $x+y=1$ lie on the straight line AB in Fig. 1, where $OA=1$ and $OB=1$. Thus this segment of the straight line AB gives a pictorial representation of the properties of the relation under the restriction to positive numbers.

Another example of a relation between two variables is afforded by considering the variations in the pressure and volume of a given mass of some gaseous substance—such as air or coal-gas or steam—at a constant temperature. Let v be the number of cubic feet in its volume and p its pressure in lb. weight per square inch. Then the law, known as Boyle's law, expressing the relation between p and v as both vary, is that the product pv is constant, always supposing that the temperature does not alter. Let us suppose, for example, that the quantity of the gas and its other circumstances are such that we can put $pv=1$ (the exact number on the right-hand side of the equation makes no essential difference).

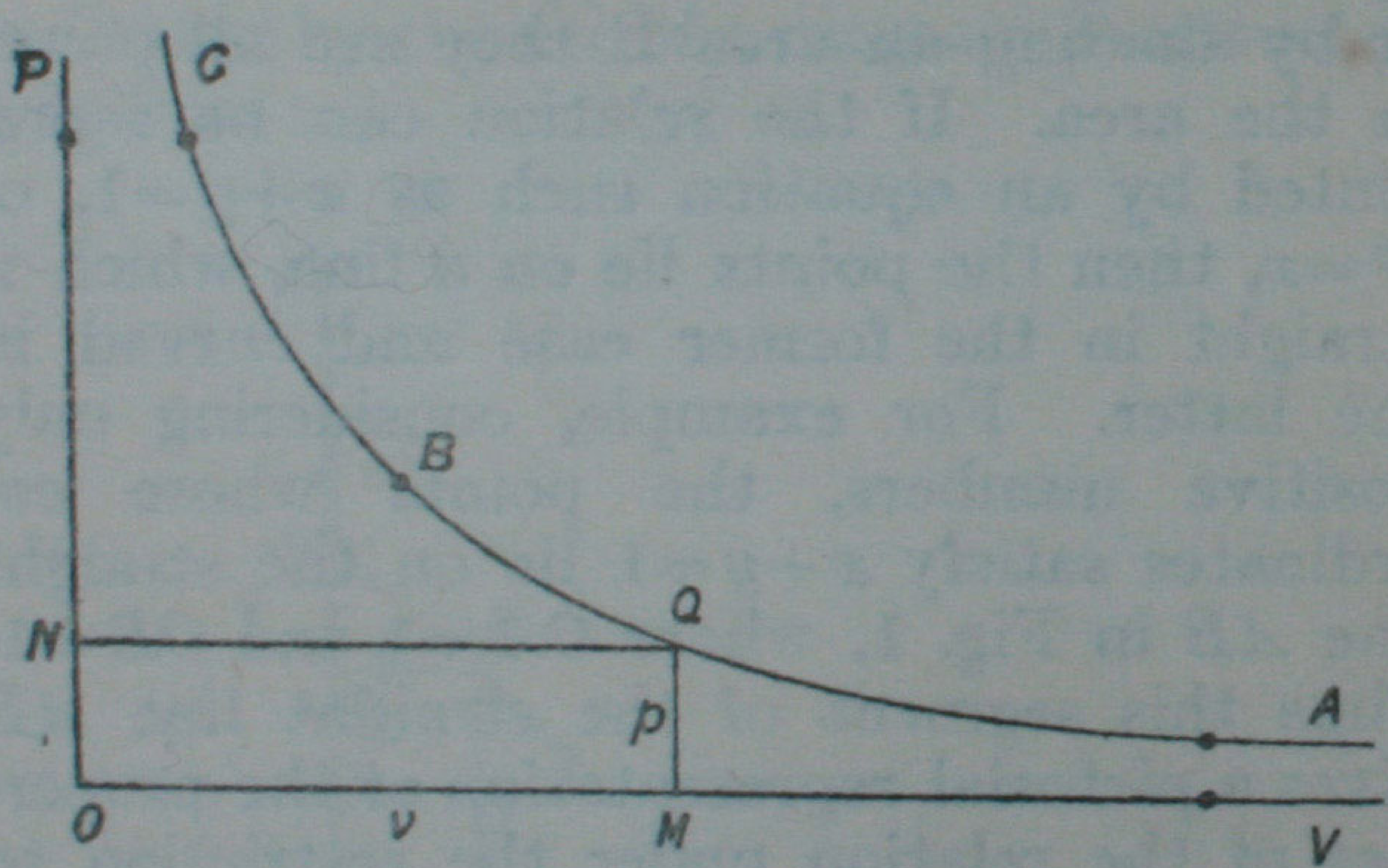


Fig. 2.

Then in Fig. 2 we take two lines, *OV* and *OP*, at right angles and draw *OM* along *OV* to represent v units of volume, and *ON* along

OP to represent p units of pressure. Then the point Q, which is found by completing the parallelogram *OMQN*, represents the state of the gas when its volume is v cubic feet and its pressure is p lb. weight per square inch. If the circumstances of the portion of gas considered are such that $pv=1$, then all these points Q which correspond to any possible state of this portion of gas must lie on the curved line *ABC*, which includes all points for which p and v are positive, and $pv=1$. Thus this curved line gives a pictorial representation of the relation holding between the volume and the pressure. When the pressure is very big the corresponding point Q must be near C, or even beyond C on the undrawn part of the curve; then the volume will be very small. When the volume is big Q will be near to A, or beyond A; and then the pressure will be small. Notice that an engineer or a physicist may want to know the particular pressure corresponding to some definitely assigned volume. Then we have the case of determining the *unknown* p when v is a known number. But this is only in particular cases. In considering generally the properties of the gas and how it will behave, he has to have in his mind the general form of the whole curve *ABC* and its general properties. In other words the really fundamental idea is that of the pair of *variables* satisfying the relation $pv = 1$. This example illustrates how the idea of *variables* is fundamental, both in the applications as well as in the theory of mathematics.

CHAPTER III

METHODS OF APPLICATION

THE way in which the idea of variables satisfying a relation occurs in the applications of mathematics is worth thought, and by devoting some time to it we shall clear up our thoughts on the whole subject.

Let us start with the simplest of examples:—Suppose that building costs 1*s.* per cubic foot and that 20*s.* make £1. Then in all the complex circumstances which attend the building of a new house, amid all the various sensations and emotions of the owner, the architect, the builder, the workmen, and the onlookers as the house has grown to completion, this fixed correlation is by the law assumed to hold between the cubic content and the cost to the owner, namely that if x be the number of cubic feet, and £y the cost, then $20y=x$. This correlation of x and y is assumed to be true for the building of any house by any owner. Also, the volume of the house and the cost are not supposed to have been perceived or apprehended by any particular sensation or faculty, or by any particular man. They are stated in an abstract general way, with complete indifference to the owner's state of mind when he has to pay the bill.

Now think a bit further as to what all this means. The building of a house is a complicated set of circumstances. It is impossible to begin to apply the law, or to test it, unless amid the general course of events it is possible to recognize a definite set of occurrences as forming a particular instance of the building of a house. In short, we must know a house when we see it, and must recognize the events which belong to its building. Then amidst these events, thus isolated in idea from the rest of nature, the two elements of the cost and cubic content must be determinable; and when they are both determined, if the law be true, they satisfy the general formula

$$20y=x.$$

But is the law true? Anyone who has had much to do with building will know that we have here put the cost rather high. It is only for an expensive type of house that it will work out at this price. This brings out another point which must be made clear. While we are making mathematical calculations connected with the formula $20y=x$, it is indifferent to us whether the law be true or false. In fact, the very meanings assigned to x and y, as being a number of cubic feet and a number of pounds sterling, are indifferent. During the mathematical investigation we are, in fact, merely considering the properties of this correlation between a pair of variable numbers x and y. Our results will apply equally well, if we interpret y to mean a number of fishermen and x the number of fish caught, so that the assumed law is that on the average each fisherman catches twenty fish. The mathematical certainty of the investigation only attaches to the results considered as giving properties of the correlation $20y=x$ between the variable pair of numbers x and y. There is no mathematical certainty whatever about the cost of the actual building of any house. The law is not quite true and the result it gives will not be quite accurate. In fact, it may well be hopelessly wrong.

Now all this no doubt seems very obvious. But in truth with more complicated instances there is no more common error than to assume that, because prolonged and accurate mathematical calculations have been made, the application of the result to some fact of nature is absolutely certain. The conclusion of no argument can be more certain than the assumptions from which it starts. All mathematical calculations about the course of nature must start from some assumed law of nature, such, for instance, as the assumed law of the cost of building stated above. Accordingly, however accurately we have calculated that some event must occur, the doubt always remains—Is the law true? If the law states a precise result, almost certainly it is not precisely accurate; and thus even at the best the result, precisely as calculated, is not likely to occur. But then we have no faculty capable of observation with ideal precision, so, after all, our inaccurate laws may be good enough.

We will now turn to an actual case, that of Newton and the Law of Gravity. This law states that any two bodies attract one another with a force proportional to the product of their masses, and inversely proportional to the square of the distance between them. Thus if m and M are the masses of the two bodies, reckoned in lbs. say, and d miles is the distance between them, the force on either body, due to the attraction of the other and directed towards it, is proportional to $\frac{mM}{d^2}$; thus this force can be written as equal to $\frac{kmM}{d^2}$, where k is a definite number depending on the absolute magnitude of this attraction and also on the scale by which we choose to measure forces. It is easy to see that, if we wish to reckon in terms of forces such as the weight of a mass of 1 lb., the number which k represents must be extremely small; for when m and M and d are each put equal to 1, $\frac{kmM}{d^2}$ becomes the gravitational attraction of two equal masses of 1 lb. at the distance of one mile, and this is quite inappreciable.

However, we have now got our formula for the force of attraction. If we call this force F, it is $F=k\frac{mM}{d^2}$, giving the correlation between the variables F, m, M, and d. We all know the story of how it was found out. Newton, it states, was sitting in an orchard and watched the fall of an apple, and then the law of universal gravitation burst upon his mind. It may be that the final formulation of the law occurred to him in an orchard, as well as elsewhere—and he must have been somewhere. But for our purposes it is more instructive to dwell upon the vast amount of preparatory thought, the product of many minds and many centuries, which was necessary before this exact law could be formulated. In the first place, the mathematical habit of mind and the mathematical procedure explained in the previous two chapters had to be generated; otherwise Newton could never have thought of a formula representing the force between *any* two masses at *any* distance. Again, what are the meanings of the terms employed, Force, Mass, Distance? Take the easiest of these terms, Distance. It seems very obvious to us to conceive all material things as forming a definite geometrical whole, such that the distances of the various parts are measurable in terms of some unit length, such as a mile or a yard. This is almost the first aspect of a material structure which occurs to us. It is the gradual outcome of the study of geometry and of the theory of measurement. Even now, in certain cases, other modes of thought are convenient. In a mountainous country distances are often reckoned in hours. But leaving distance, the other terms, Force and Mass, are much more obscure. The exact comprehension of the ideas which Newton meant to convey by these words was of slow growth, and, indeed, Newton himself was the first man who had thoroughly mastered the true general principles of Dynamics.

Throughout the middle ages, under the influence of Aristotle, the science was entirely misconceived. Newton had the advantage of coming after a series of great men, notably Galileo, in Italy, who in the previous two centuries had reconstructed the science and had invented the right way of thinking about it. He completed their work. Then, finally, having the ideas of force, mass, and distance, clear and distinct in his mind, and realising their importance and their relevance to the fall of an apple and the motions of the planets, he hit upon the law of gravitation and proved it to be the formula always satisfied in these various motions.

The vital point in the application of mathematical formulæ is to have clear ideas and a correct estimate of their relevance to the phenomena under observation. No less than ourselves, our remote ancestors were impressed with the importance of natural phenomena and with the desirability of taking energetic measures to regulate the sequence of events. Under the influence of irrelevant ideas they executed elaborate religious ceremonies to aid the birth of the new moon, and performed sacrifices to save the sun during the crisis of an eclipse. There is no reason to believe that they were more stupid than we are. But at that epoch there had not been opportunity for the slow accumulation of clear and relevant ideas.

The sort of way in which physical sciences grow into a form capable of treatment by mathematical methods is illustrated by the history of the gradual growth of the science of electromagnetism. Thunderstorms are events on a grand scale, arousing terror in men and even animals. From the earliest times they must have been objects of wild and fantastic hypotheses, though it may be doubted whether our modern scientific discoveries in connection with electricity are not more astonishing than any of the magical explanations of savages. The Greeks knew that amber (Greek, electron) when rubbed would attract light and dry bodies. In 1600 A.D., Dr. Gilbert, of Colchester, published the first work on the subject in which any scientific method is followed. He made a list of substances possessing properties similar to those of amber; he must also have the credit of connecting, however vaguely, electric and magnetic phenomena. At the end of the seventeenth and throughout the eighteenth century knowledge advanced. Electrical machines were made, sparks were obtained from them; and the Leyden Jar was invented, by which these effects could be intensified. Some organised knowledge was being obtained; but still no relevant mathematical ideas had been found out. Franklin, in the year 1752, sent a kite into the clouds and proved that thunderstorms were electrical.

Meanwhile from the earliest epoch (2634 B.C.) the Chinese had utilized the characteristic property of the compass needle, but do not seem to have connected it with any theoretical ideas. The really profound changes in human life all have their ultimate origin in knowledge pursued for its own sake. The use of the compass was not introduced into Europe till the end of the twelfth century A.D., more than 3000 years after its first use in China. The importance which the science of electromagnetism has since assumed in every department of human life is not due to the superior practical bias of Europeans, but to the fact that in the West electrical and magnetic phenomena were studied by men who were dominated by abstract theoretic interests.

The discovery of the electric current is due to two Italians, Galvani in 1780, and Volta in 1792. This great invention opened a new series of phenomena for investigation. The scientific world had now three separate, though allied, groups of occurrences on hand—the effects of "statical" electricity arising from frictional electrical machines, the magnetic phenomena, and the effects due to electric currents. From the end of the eighteenth century onwards, these three lines of investigation were quickly inter-connected and the modern science of electromagnetism was constructed, which now threatens to transform human life.

Mathematical ideas now appear. During the decade 1780 to 1789, Coulomb, a Frenchman, proved that magnetic poles attract or repel each other, in proportion to the inverse square of their distances, and also that the same law holds for electric charges—laws curiously analogous to that of gravitation. In 1820, Öersted, a Dane, discovered that electric currents exert a force on magnets, and almost immediately afterwards the mathematical law of the force was correctly formulated by Ampère, a Frenchman, who also proved that two electric currents exerted forces on each other. "The experimental investigation by which Ampère established the law of the mechanical action between electric currents is one of the most brilliant achievements in science. The whole, theory and experiment, seems as if it had leaped, full-grown and full armed, from the brain of the 'Newton of Electricity.' It is perfect in form, and unassailable in accuracy, and it is summed up in a formula from which all the phenomena may be deduced, and which must always remain the cardinal formula of electro-dynamics." *

The momentous laws of induction between currents and between currents and magnets were discovered by Michael Faraday in 1831–32. Faraday was asked: "What is the use of this discovery?" He answered: "What is the use of a child—it grows to be a man." Faraday's child has grown to be a man and is now the basis of all the modern applications

* *Electricity and Magnetism*, Clerk Maxwell, Vol. II., ch. iii.

of electricity. Faraday also reorganized the whole theoretical conception of the science. His ideas, which had not been fully understood by the scientific world, were extended and put into a directly mathematical form by Clerk Maxwell in 1873. As a result of his mathematical investigations, Maxwell recognized that, under certain conditions, electrical vibrations ought to be propagated. He at once suggested that the vibrations which form light are electrical. This suggestion has since been verified, so that now the whole theory of light is nothing but a branch of the great science of electricity. Also Herz, a German, in 1888, following on Maxwell's ideas, succeeded in producing electric vibrations by direct electrical methods His experiments are the basis of our wireless telegraphy.

In more recent years even more fundamental discoveries have been made, and the science continues to grow in theoretic importance and in practical interest. This rapid sketch of its progress illustrates how, by the gradual introduction of the relevant theoretic ideas, suggested by experiment and themselves suggesting fresh experiments, a whole mass of isolated and even trivial phenomena are welded together into one coherent science, in which the results of abstract mathematical deductions, starting from a few simple assumed laws, supply the explanation to the complex tangle of the course of events.

Finally, passing beyond the particular sciences of electromagnetism and light, we can generalize our point of view still further, and direct our attention to the growth of mathematical physics considered as one great chapter of scientific thought. In the first place, what in the barest outlines is the story of its growth ?

It did not begin as one science, or as the product of one band of men. The Chaldean shepherds watched the skies, the agents of Government in Mesopotamia and Egypt measured the land, priests and philosophers brooded on the general nature of all things. The vast mass of the operations of nature appeared due to mysterious unfathomable forces. " The wind bloweth where it listeth " expresses accurately the blank ignorance then existing of any stable rules followed in detail by the succession of phenomena. In broad outline, then as now, a regularity of events was patent. But no minute tracing of their interconnection was possible, and there was no knowledge how even to set about to construct such a science.

Detached speculations, a few happy or unhappy shots at the nature of things, formed the utmost which could be produced.

Meanwhile land-surveys had produced geometry, and the observations of the heavens disclosed the exact regularity of the solar system. Some of the later Greeks, such as Archimedes, had just views on the elementary phenomena of hydrostatics and optics. Indeed, Archimedes, who combined a genius for mathematics with a physical insight, must rank with Newton, who lived nearly two thousand years later, as one of the founders of mathematical physics. He lived at Syracuse, the great Greek city of Sicily. When the Romans besieged the town (in 212 to 210 B.C.), he is said to have burned their ships by concentrating on them, by means of mirrors, the sun's rays. The story is highly improbable, but is good evidence of the reputation which he had gained among his contemporaries for his knowledge of optics. At the end of this siege he was killed. According to one account given by Plutarch, in his life of Marcellus, he was found by a Roman soldier absorbed in the study of a geometrical diagram which he had traced on the sandy floor of his room. He did not immediately obey the orders of his captor, and so was killed. For the credit of the Roman generals it must be said that the soldiers had orders to spare him. The internal evidence for the other famous story of him is very strong; for the discovery attributed to him is one eminently worthy of his genius for mathematical and physical research. Luckily, it is simple enough to be explained here in detail. It is one of the best easy examples of the method of application of mathematical ideas to physics.

Hiero, King of Syracuse, had sent a quantity of gold to some goldsmith to form the material of a crown. He suspected that the craftsmen had abstracted some of the gold and had supplied its place by alloying the remainder with some baser metal. Hiero sent the crown to Archimedes and asked him to test it. In these days an indefinite number of chemical tests would be available. But then Archimedes had to think out the matter afresh. The solution flashed upon him as he lay in his bath. He jumped up and ran through the streets to the palace, shouting *Eureka! Eureka!* (I have found it, I have found it). This day, if we knew which it was, ought to be celebrated as the birthday of mathematical physics; the science came of age when Newton sat in his orchard. Archimedes had in truth made a great discovery. He saw that a body when immersed in water is pressed upwards by the surrounding water with a resultant force equal to the weight of the water it displaces. This law can be proved theoretically from the mathematical principles of hydrostatics and can also be verified experimentally. Hence, if W lb. be the weight of the crown, as weighed in air, and w lb. be the weight of the water which it displaces when completely immersed, $W-w$ would be the extra upward force necessary to sustain the crown as it hung in water.

Now, this upward force can easily be ascertained by weighing the body as it hangs in water, as shown in the annexed figure. If

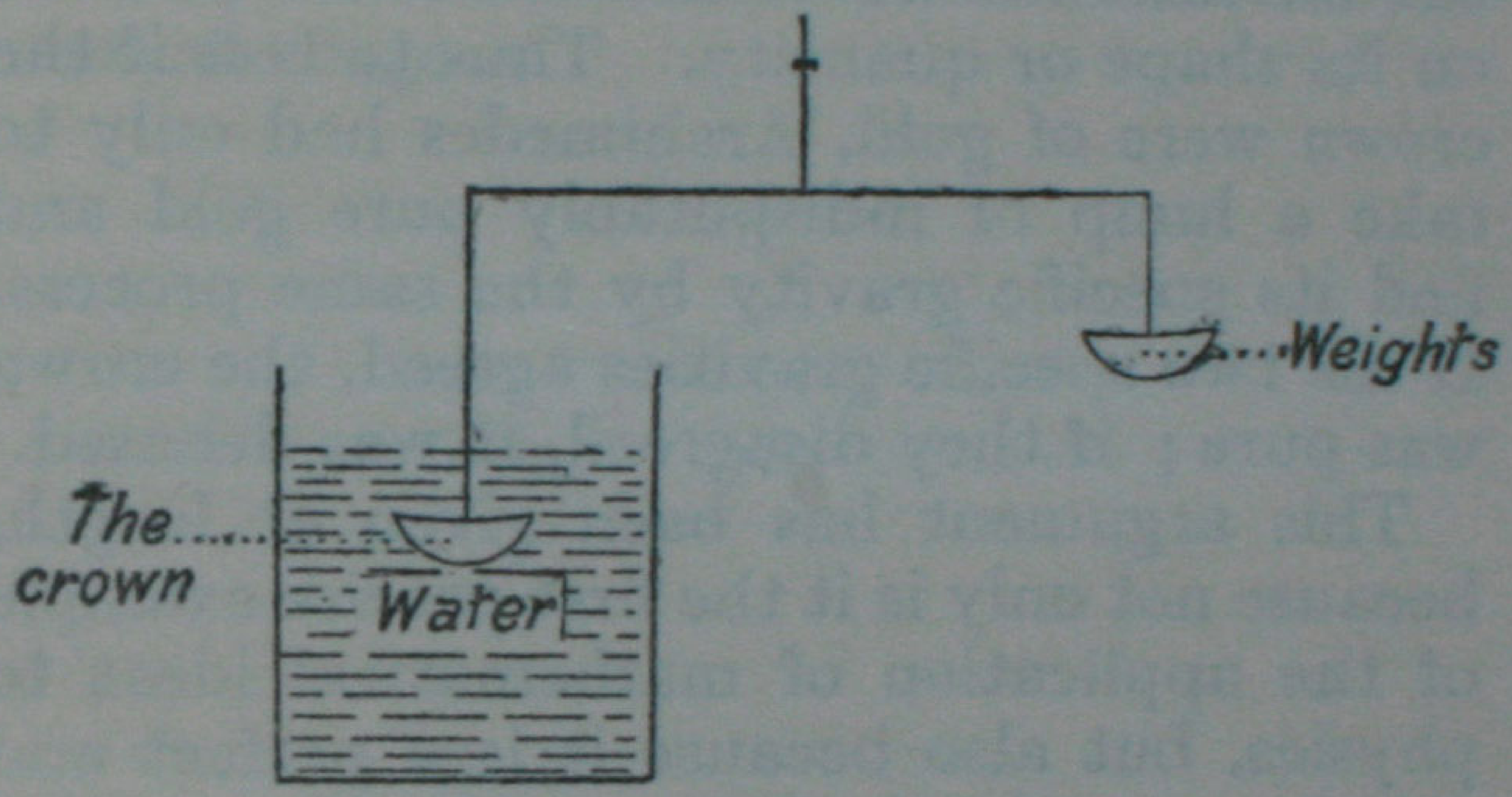


Fig. 3.

the weights in the right-hand scale come to F lb., then the apparent weight of the crown in water is F lb.; and we thus have

$$F = W - w$$

and thus $$w = W - F,$$

and $$\frac{W}{w} = \frac{W}{W-F} \qquad (A)$$

where W and F are determined by the easy, and fairly precise, operation of weighing.

Hence, by equation (A), $\frac{W}{w}$ is known. But $\frac{W}{w}$ is the ratio of the weight of the crown to the weight of an equal volume of water. This ratio is the same for any lump of metal of the same material: it is now called the specific gravity of the material, and depends only on the intrinsic nature of the substance and not on its shape or quantity. Thus to test if the crown were of gold, Archimedes had only to take a lump of indisputably pure gold and find its specific gravity by the same process. If the two specific gravities agreed, the crown was pure; if they disagreed, it was debased.

This argument has been given at length, because not only is it the first precise example of the application of mathematical ideas to physics, but also because it is a perfect and simple example of what must be the method and spirit of the science for all time.

The death of Archimedes by the hands of a Roman soldier is symbolical of a world-change of the first magnitude: the theoretical Greeks, with their love of abstract science, were superseded in the leadership of the European world by the practical Romans. Lord Beaconsfield, in one of his novels, has defined a practical man as a man who practises the errors of his forefathers. The Romans were a great race, but they were cursed with the sterility which waits upon practicality. They did not improve upon the knowledge of their forefathers, and all their advances were confined to the minor technical details of engineering. They were not dreamers enough to arrive at new points of view, which could give a more fundamental control over the forces of nature. No Roman lost his life because he was absorbed in the contemplation of a mathematical diagram.

CHAPTER IV

DYNAMICS

The world had to wait for eighteen hundred years till the Greek mathematical physicists found successors. In the sixteenth and seventeenth centuries of our era great Italians, in particular Leonardo da Vinci, the artist (born 1452, died 1519), and Galileo (born 1564, died 1642), rediscovered the secret, known to Archimedes, of relating abstract mathematical ideas with the experimental investigation of natural phenomena. Meanwhile the slow advance of mathematics and the accumulation of accurate astronomical knowledge had placed natural philosophers in a much more advantageous position for research. Also the very egoistic self-assertion of that age, its greediness for personal experience, led its thinkers to want to see for themselves what happened; and the secret of the relation of mathematical theory and experiment in inductive reasoning was practically discovered. It was an act eminently characteristic of the age that Galileo, a philosopher, should have dropped the weights from the leaning tower of Pisa. There are always men of thought and men of action; mathematical physics is the product of an age which combined in the same men impulses to thought with impulses to action.

This matter of the dropping of weights from the tower marks picturesquely an essential step in knowledge, no less a step than the first attainment of correct ideas on the science of dynamics, the basal science of the whole subject. The particular point in dispute was as to whether bodies of different weights would fall from the same height in the same time. According to a dictum of Aristotle, universally followed up to that epoch, the heavier weight would fall the quicker. Galileo affirmed that they would fall in the same time, and proved his point by dropping weights from the top of the leaning tower. The apparent exceptions to the rule all arise when, for some reason, such as extreme lightness or great speed, the air resistance is important. But neglecting the air the law is exact.

Galileo's successful experiment was not the result of a mere lucky guess. It arose from his correct ideas in connection with inertia and mass. The first law of motion, as following Newton we now enunciate it, is—Every body continues in its state of rest or of uniform motion in a straight line, except so far as it is compelled by impressed force to change that state. This law is more than a dry formula: it is also a pæan of triumph over defeated heretics. The point at issue can be understood by deleting from the law the phrase "or of uniform motion in a straight line." We there obtain what might be taken as the Aristotelian opposition formula: "Every body continues in its state of rest except so far as it is compelled by impressed force to change that state."

In this last false formula it is asserted that, apart from force, a body continues in a state of rest; and accordingly that, if a body is moving, a force is required to sustain the motion; so that when the force ceases, the motion ceases. The true Newtonian law takes diametrically the opposite point of view. The state of a body unacted on by force is that of uniform motion in a straight line, and no external force or influence is to be looked for as the cause, or, if you like to put it so, as the invariable accompaniment of this uniform rectilinear motion. Rest is merely a particular case of such motion, merely when the velocity is and remains zero. Thus, when a body is moving, we do not seek for any external influence except to explain changes in the rate of the velocity or changes in its direction. So long as the body is moving at the same rate and in the same direction there is no need to invoke the aid of any forces.

The difference between the two points of view is well seen by reference to the theory of the motion of the planets. Copernicus, a Pole, born at Thorn in West Prussia (born 1473, died 1543), showed how much simpler it was to conceive the planets, including the

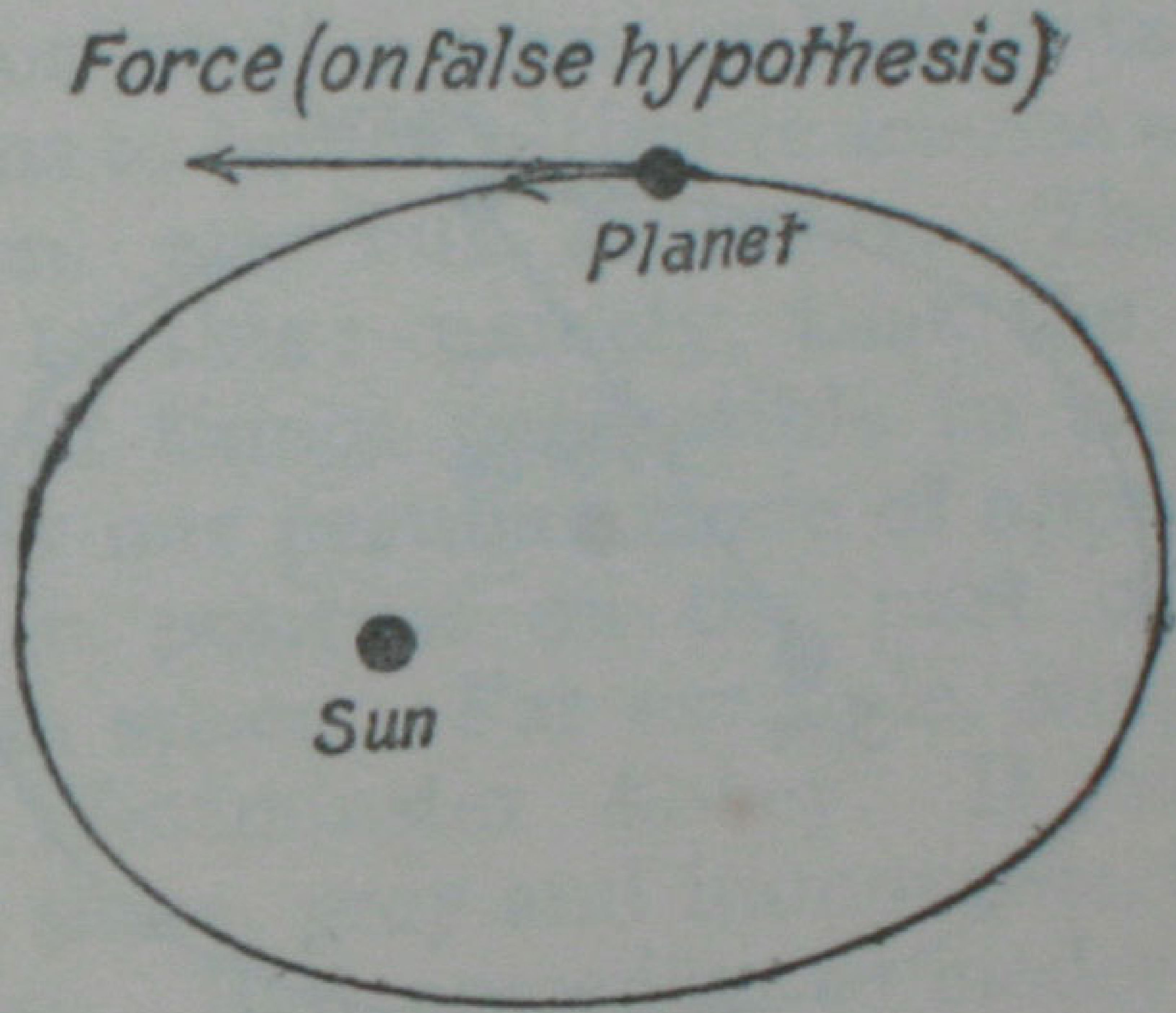


Fig. 4.

earth as revolving round the sun in orbits which are nearly circular; and later, Kepler, a German mathematician, in the year 1609 proved that, in fact, the orbits are practically ellipses, that is, a special sort of oval curves which we will consider later in more detail. Immediately the question arose as to what are the forces which preserve the planets in this motion. According to the old false view, held by Kepler, the actual velocity itself required preservation by force. Thus he looked for tangential forces as in the accompanying figure (4). But according to the Newtonian law, apart from some force the planet would move for ever with its existing velocity in a straight line, and thus depart entirely from the sun. Newton, therefore, had to search for a force which would bend the motion

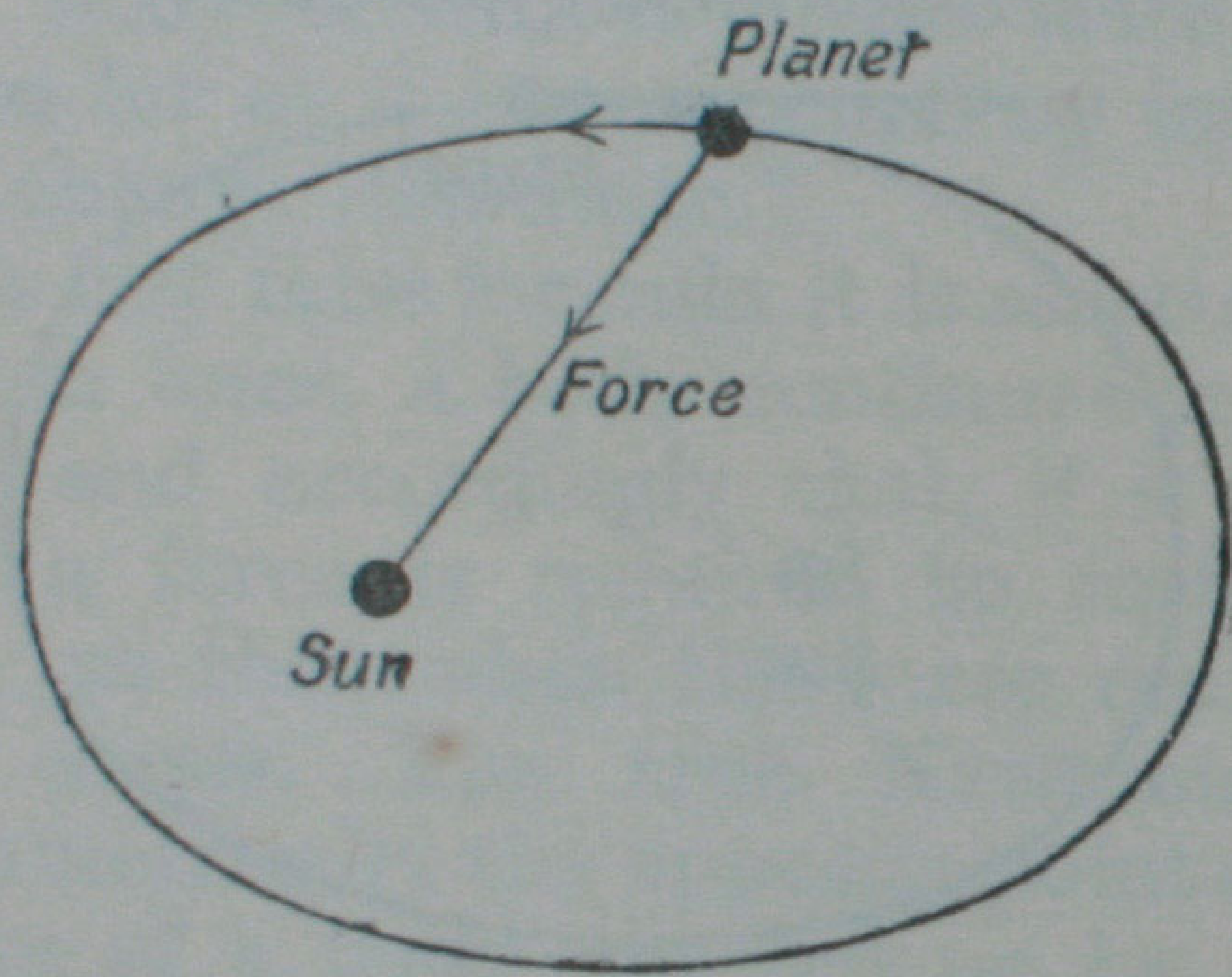


Fig. 5.

round into its elliptical orbit. This he showed must be a force directed towards the sun as in the next figure (5). In fact, the force is the gravitational attraction of the sun acting according to the law of the inverse square of the distance, which has been stated above.

The science of mechanics rose among the Greeks from a consideration of the theory of the mechanical advantage obtained by the use of a lever, and also from a consideration of various problems connected with the weights of bodies. It was finally put on its true basis at the end of the sixteenth and during the seventeenth centuries, as the preceding account shows, partly with the view of explaining the theory of falling bodies, but chiefly in order to give a scientific theory of planetary motions. But since those days dynamics has taken upon itself a more ambitious task, and now claims to be the ultimate science of which the others are but branches. The claim amounts to this: namely, that the various qualities of things perceptible to the senses are merely our peculiar mode of appreciating changes in position on the part of things existing in space. For example, suppose we look at Westminster Abbey. It has been standing there, grey and immovable, for centuries past. But, according to modern scientific theory, that greyness, which so heightens our sense of the immobility of the building, is itself nothing but our way of appreciating the rapid motions of the ultimate molecules, which form the outer surface of the building and communicate vibrations to a substance called the ether. Again we lay our hands on its stones and note their cool, even temperature, so symbolic of the quiet repose of the building. But this feeling of temperature simply marks our sense of the transfer of heat from the hand to the stone, or from the stone to the hand; and, according to modern science, heat is nothing but the agitation of the molecules of a body. Finally, the organ begins playing, and again sound is nothing but the result of motions of the air striking on the drum of the ear.

Thus the endeavour to give a dynamical explanation of phenomena is the attempt to explain them by statements of the general form, that such and such a substance or body was in this place and is now in that place. Thus we arrive at the great basal idea of modern science, that all our sensations are the result of comparisons of the changed configurations of things in space at various times. It follows therefore, that the laws of motion, that is, the laws of the changes of configurations of things, are the ultimate laws of physical science.

In the application of mathematics to the investigation of natural philosophy, science does systematically what ordinary thought does casually. When we talk of a chair, we usually mean something which we have been seeing or feeling in some way; though most of our language will presuppose that there is something which exists independently of our sight or feeling. Now in mathematical physics the opposite course is taken. The chair is conceived without any reference to anyone in particular, or to any special modes of perception. The result is that the chair becomes in thought a set of molecules in space, or a group of electrons, a portion of the ether in motion, or however the current scientific ideas describe it. But the point is that science reduces the chair to things moving in space and influencing each other's motions. Then the various elements or factors which enter into a set of circumstances, as thus conceived, are merely the things, like lengths of lines, sizes of angles, areas, and volumes, by which the positions of bodies in space can be settled. Of course, in addition to these geometrical elements the fact of motion and change necessitates the introduction of the rates of changes of such elements, that is to say, velocities, angular velocities, accelerations, and suchlike things. Accordingly, mathematical physics deals with correlations between variable numbers which are supposed to represent the correlations which exist in nature between the measures of these geometrical elements and of their rates of change. But always the mathematical laws deal with variables, and it is only in the occasional testing of the laws by reference to experiments, or in the use of the laws for special predictions that definite numbers are substituted.

The interesting point about the world as thus conceived in this abstract way throughout the study of mathematical physics, where only the positions and shapes of things are considered together with their changes, is that the events of such an abstract world are sufficient to "explain" our sensations. When we hear a sound, the molecules of the air have been agitated in a certain way: given the agitation, or air-waves as they are called, all normal people hear sound; and if there are no air-waves, there is no sound. And, similarly, a physical cause or origin, or parallel event (according as different people might like to phrase it) underlies our other sensations. Our very thoughts appear to correspond to conformations and motions of the brain; injure the brain and you injure the thoughts. Meanwhile the events of this physical universe succeed each other according to the mathematical laws which ignore all special sensations and thoughts and emotions.

Now, undoubtedly, this is the general aspect of the relation of the world of mathematical physics to our emotions, sensations, and thoughts; and a great deal of controversy has been occasioned by it and much ink spilled. We need only make one remark. The whole situation has arisen, as we have seen, from the endeavour to describe an external world "explanatory" of our various individual sensations and emotions, but a world also, not essentially dependent upon any particular sensations or upon any particular individual. Is such a world merely but one huge fairy tale? But fairy tales are fantastic and arbitrary: if in truth there be such a world, it ought to submit itself to an exact description, which determines accurately its various parts and their mutual relations. Now, to a large degree, this scientific world does submit itself to this test and allow its events to be explored and predicted by the apparatus of abstract mathematical ideas. It certainly seems that here we have an inductive verification of our initial assumption. It must be admitted that no inductive proof is conclusive; but if the whole idea of a world which has existence independently of our particular perceptions of it be erroneous, it requires careful explanation why the attempt to characterise it, in terms of that mathematical remnant of our ideas which would apply to it, should issue in such a remarkable success.

It would take us too far afield to enter into a detailed explanation of the other laws of motion. The remainder of this chapter must be devoted to the explanation of remarkable ideas which are fundamental, both to mathematical physics and to pure mathematics: these are the ideas of vector quantities and the parallelogram law for vector addition. We have seen that the essence of motion is that a body was at A and is now at C. This transference from A to C requires two distinct elements to be settled before it is completely determined, namely its magnitude (*i.e.* the length AC) and its direction. Now anything, like this transference, which is completely given by the determination of a magni-

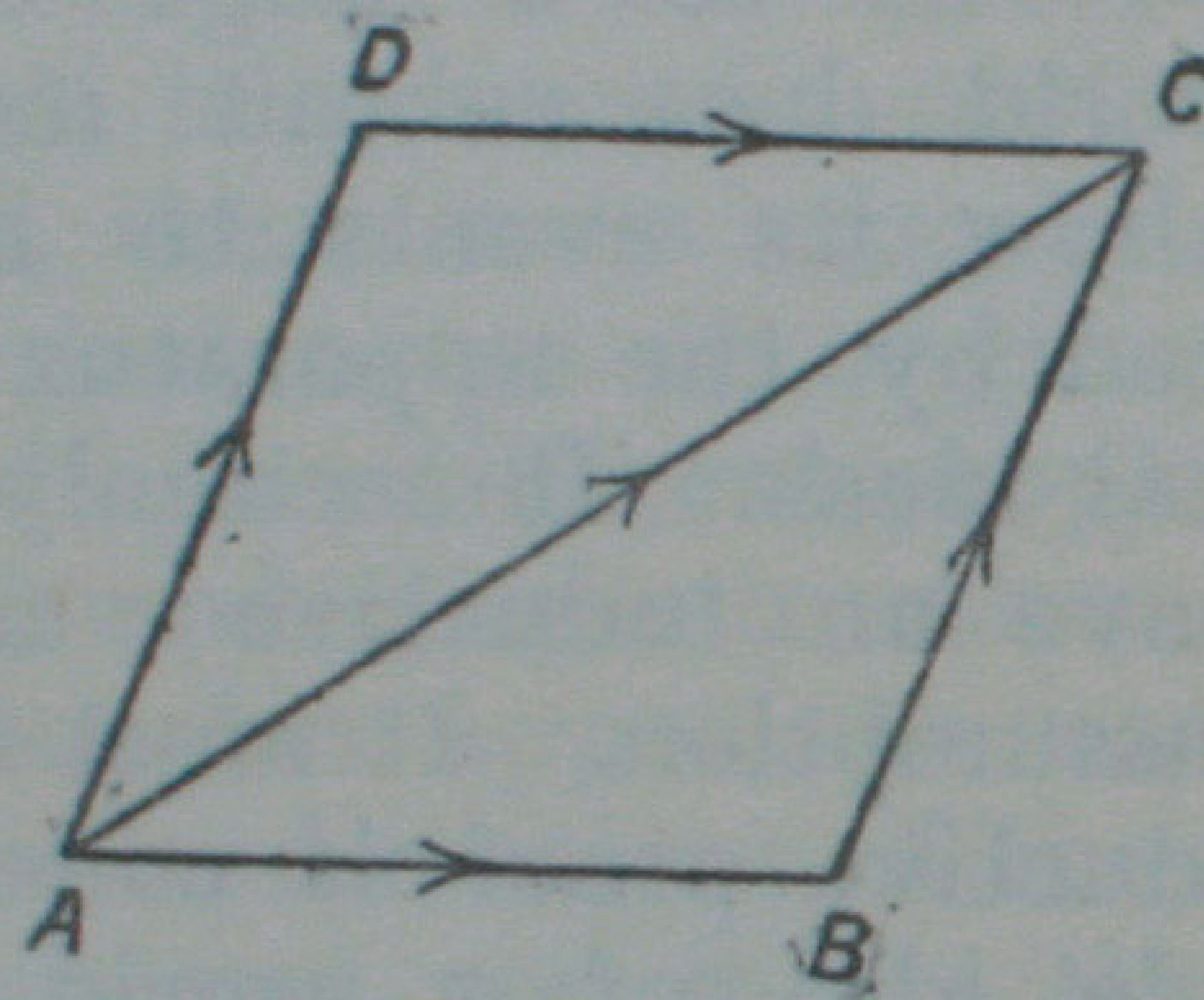


Fig. 6.

tude and a direction is called a vector. For example, a velocity requires for its definition the assignment of a magnitude and of a direction. It must be of so many miles per hour in such and such a direction. The existence and the independence of these two elements in the determination of a velocity are well illustrated by the action of the captain of a ship, who communicates with different subordinates respecting them: he tells the chief engineer the number of knots at which he is to steam, and the helmsman the compass bearing of the course which he is to keep. Again the rate of change of velocity, that is velocity added per unit time, is also a vector quantity: it is called the acceleration. Similarly a force in the dynamical sense is another vector quantity. Indeed, the vector nature of forces follows at once according to dynamical principles from that of velocities and accelerations; but this is a point which we need not go into. It is sufficient here to say that a force acts on a body with a certain magnitude in a certain direction.

Now all vectors can be graphically represented by straight lines. All that has to be done is to arrange: (i) a scale according to which units of length correspond to units of magnitude of the vector—for example, one inch to a velocity of 10 miles per hour in the case of velocities, and one inch to a force of 10 tons weight in the case of forces—and (ii) a direction of the line on the diagram corresponding to the direction of the vector. Then a line drawn with the proper number of inches of length in the proper direction represents the required vector on the arbitrarily assigned scale of magnitude. This diagrammatic representation of vectors is of the first importance. By its aid we can enunciate the famous "parallelogram law" for the addition of vectors of the same kind but in different directions.

Consider the vector *AC* in figure 6 as representative of the changed position of a body from A to C: we will call this the vector of transportation. It will be noted that, if the reduction of physical phenomena to mere changes in positions, as explained above, is correct, all other types of physical vectors are really reducible in some way or other to this single type. Now the final transportation from A to C is equally well effected by a transportation from A to B and a transportation from B to C, or, completing the parallelogram $ABCD$, by a transportation from A to D and a transportation from D to C. These transportations as thus successively applied are said to be added together. This is simply a definition of what we mean by the addition of transportations. Note further that, considering parallel lines as being lines drawn in the same direction, the transportations B to C and A to D may be conceived as the same transportation applied to bodies in the two initial positions B and A. With this conception we may talk of the transportation A to D as applied to a body in any position, for example at B. Thus we may say that the transportation A to C can be conceived as the sum of the two transportations A to B and A to D applied in any order. Here we have the parallelogram law for the addition of transportations: namely, if the transportations are A to B and A to D, complete the parallelogram *ABCD*, and then the sum of the two is the diagonal *AC*.

All this at first sight may seem to be very artificial. But it must be observed that nature itself presents us with the idea. For example, a steamer is moving in the direction *AD* (cf. fig. 6) and a man walks across its deck. If the steamer were still, in one minute he would arrive at *B*; but during that minute his starting point *A* on the deck has moved to *D*, and his path on the deck has moved from *AB* to *DC*. So that, in fact, his transportation has been from *A* to *C* over the surface of the sea. It is, however, presented to us analysed into the sum of two transportations, namely, one from *A* to *B* relatively to the steamer, and one from *A* to *D* which is the transportation of the steamer.

By taking into account the element of time, namely one minute, this diagram of the man's transportation *AC* represents his velocity. For if *AC* represented so many feet of transportation, it now represents a transportation of so many feet per minute, that is to say, it represents the velocity of the man. Then *AB* and *AD* represent two velocities, namely, his velocity relatively to the steamer, and the velocity of the steamer, whose "sum" makes up his complete velocity. It is evident that diagrams and definitions concerning transportations are turned into diagrams and definitions concerning velocities by conceiving the diagrams as representing transportations per unit time. Again, diagrams and definitions concerning velocities are turned into diagrams and definitions concerning accelera-

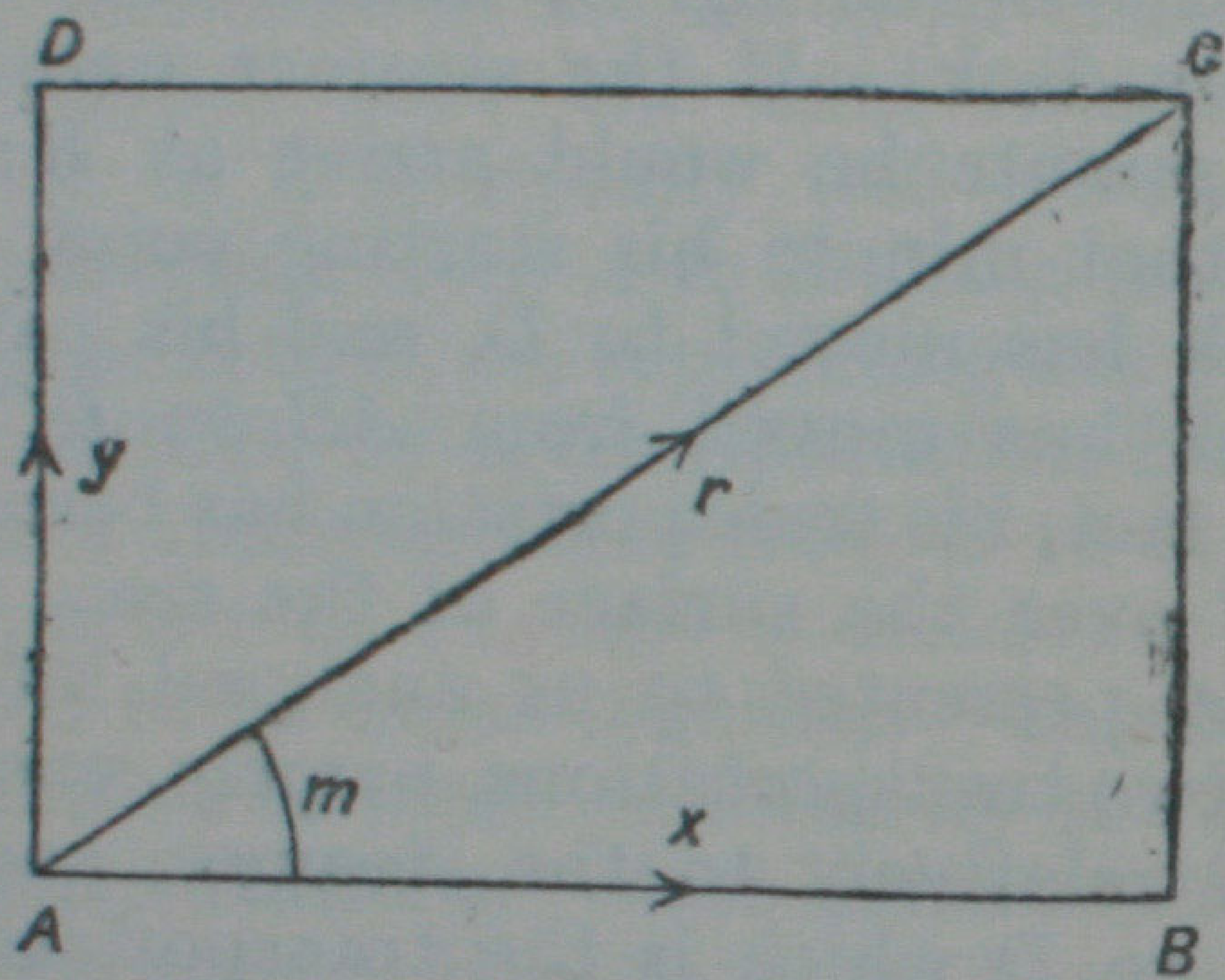


Fig. 7.

tions by conceiving the diagrams as representing velocities added per unit time.

Thus by the addition of vector velocities and of vector accelerations, we mean the addition according to the parallelogram law.

Also, according to the laws of motion a force is fully represented by the vector acceleration it produces in a body of given mass. Accordingly, forces will be said to be added when their joint effect is to be reckoned according to the parallelogram law.

Hence for the fundamental vectors of

science, namely transportations, velocities, and forces, the addition of any two of the same kind is the production of a "resultant" vector according to the rule of the parallelogram law.

By far the simplest type of parallelogram is a rectangle, and in pure mathematics it is the relation of the single vector AC to the two component vectors, AB and AD, at right angles (cf. fig. 7), which is continually recurring. Let x, y, and r units represent the lengths of AB, AD, and AC, and let m units of angle represent the magnitude of the angle BAC. Then the relations between x, y, r, and m, in all their many aspects are the continually recurring topic of pure mathematics; and the results are of the type required for application to the fundamental vectors of mathematical physics. This diagram is the chief bridge over which the results of pure mathematics pass in order to obtain application to the facts of nature.

CHAPTER V

THE SYMBOLISM OF MATHEMATICS

WE now return to pure mathematics, and consider more closely the apparatus of ideas out of which the science is built. Our first concern is with the symbolism of the science, and we start with the simplest and universally known symbols, namely those of arithmetic.

Let us assume for the present that we have sufficiently clear ideas about the integral numbers, represented in the Arabic notation by 0,1,2, . . ., 9, 10, 11, . . . 100, 101, . . . and so on. This notation was introduced into Europe through the Arabs, but they apparently obtained it from Hindoo sources. The first known work * in which it is systematically explained is a work by an Indian mathematician, Bhaskara (born 1114 A.D.). But the actual numerals can be traced back to the seventh century of our era, and perhaps were originally invented in Tibet. For our present

* For the detailed historical facts relating to pure mathematics, I am chiefly indebted to *A Short History of Mathematics*, by W. W. R. Ball.

purposes, however, the history of the notation is a detail. The interesting point to notice is the admirable illustration which this numeral system affords of the enormous importance of a good notation. By relieving the brain of all unnecessary work, a good notation sets it free to concentrate on more advanced problems, and in effect increases the mental power of the race. Before the introduction of the Arabic notation, multiplication was difficult, and the division even of integers called into play the highest mathematical faculties. Probably nothing in the modern world would have more astonished a Greek mathematician than to learn that, under the influence of compulsory education, a large proportion of the population of Western Europe could perform the operation of division for the largest numbers. This fact would have seemed to him a sheer impossibility. The consequential extension of the notation to decimal fractions was not accomplished till the seventeenth century. Our modern power of easy reckoning with decimal fractions is the almost miraculous result of the gradual discovery of a perfect notation.

Mathematics is often considered a difficult and mysterious science, because of the numerous symbols which it employs. Of course, nothing is more incomprehensible than a symbolism which we do not understand. Also a symbolism, which we only partially understand and are unaccustomed to use, is difficult to follow. In exactly the same way the technical terms of any profession or trade are incomprehensible to those who have never been trained to use them. But this is not because they are difficult in themselves. On the contrary they have invariably been introduced to make things easy. So in mathematics, granted that we are giving any serious attention to mathematical ideas, the symbolism is invariably an immense simplification. It is not only of practical use, but is of great interest. For it represents an analysis of the ideas of the subject and an almost pictorial representation of their relations to each other. If anyone doubts the utility of symbols, let him write out in full, without any symbol whatever, the whole meaning of the following equations which represent some of the fundamental laws of algebra * :—

$$x+y=y+x \quad .. \quad .. \quad .. \quad (1)$$
$$(x+y)+z=x+(y+z) \quad .. \quad .. \quad (2)$$
$$x \times y=y \times x \quad .. \quad .. \quad .. \quad (3)$$
$$(x \times y) \times z=x \times (y \times z) \quad .. \quad (4)$$
$$x \times (y+z)=(x \times y)+(x \times z) \quad .. \quad (5)$$

Here (1) and (2) are called the commutative and associative laws for addition, (3) and (4)

* Cf. Note A, p. 250.

are the commutative and associative laws for multiplication, and (5) is the distributive law relating addition and multiplication. For example, without symbols, (1) becomes: If a second number be added to any given number the result is the same as if the first given number had been added to the second number.

This example shows that, by the aid of symbolism, we can make transitions in reasoning almost mechanically by the eye, which otherwise would call into play the higher faculties of the brain.

It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle—they are strictly limited in number, they require fresh horses, and must only be made at decisive moments.

One very important property for symbolism to possess is that it should be concise, so as to be visible at one glance of the eye and to be rapidly written. Now we cannot place symbols more concisely together than by placing them in immediate juxtaposition. In a good symbolism therefore, the juxtaposition of important symbols should have an important meaning. This is one of the merits of the Arabic notation for numbers; by means of ten symbols, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and by simple juxtaposition it symbolizes any number whatever. Again in algebra, when we have two variable numbers x and y, we have to make a choice as to what shall be denoted by their juxtaposition xy. Now the two most important ideas on hand are those of addition and multiplication. Mathematicians have chosen to make their symbolism more concise by defining xy to stand for $x \times y$. Thus the laws (3), (4), and (5) above are in general written,

$$xy=yx, \quad (xy)z=x(yz), \quad x(y+z)=xy+xz,$$

thus securing a great gain in conciseness. The same rule of symbolism is applied to the juxtaposition of a definite number and a variable: we write $3x$ for $3 \times x$, and $30x$ for $30 \times x$.

It is evident that in substituting definite numbers for the variables some care must be taken to restore the $\times$, so as not to conflict with the Arabic notation. Thus when we substitute 2 for x and 3 for y in xy, we must write 2×3 for xy, and not 23 which means $20+3$.

It is interesting to note how important for the development of science a modest-looking symbol may be. It may stand for the emphatic presentation of an idea, often a very subtle idea, and by its existence make it easy to exhibit the relation of this idea to all the complex trains of ideas in which it occurs. For example, take the most modest of all symbols, namely, 0, which stands for the *number* zero. The Roman notation for numbers had no symbol for zero, and probably most mathematicians of the ancient world would have been horribly puzzled by the idea of the number zero. For, after all, it is a very subtle idea, not at all obvious. A great deal of discussion on the meaning of the zero of quantity will be found in philosophic works. Zero is not, in real truth, more difficult or subtle in idea than the other cardinal numbers. What do we mean by 1 or by 2, or by 3? But we are familiar with the use of these ideas, though we should most of us be puzzled to give a clear analysis of the simpler ideas which go to form them. The point about zero is that we do not need to use it in the operations of daily life. No one goes out to buy zero fish. It is in a way the most civilized of all the cardinals, and its use is only forced on us by the needs of cultivated modes of thought. Many important services are rendered by the symbol 0, which stands for the number zero.

The symbol developed in connection with the Arabic notation for numbers of which it is an essential part. For in that notation the value of a digit depends on the position in which it occurs. Consider, for example, the digit 5, as occurring in the numbers 25, 51, 3512, 5213. In the first number 5 stands for five, in the second number 5 stands for fifty, in the third number for five hundred, and in the fourth number for five thousand. Now, when we write the number fifty-one in the symbolic form 51, the digit 1 pushes the digit 5 along to the second place (reckoning from right to left) and thus gives it the value fifty. But when we want to symbolize fifty by itself, we can have no digit 1 to perform this service; we want a digit in the units place to add nothing to the total and yet to push the 5 along to the second place. This service is performed by 0, the symbol for zero. It is extremely probable that the men who introduced 0 for this purpose had no definite conception in their minds of the number zero. They simply wanted a mark to symbolize the fact that nothing was contributed by the digit's place in which it occurs. The idea of zero probably took shape gradually from a desire to assimilate the meaning of this mark to that of the marks, 1, 2, . . . 9, which do represent cardinal numbers. This would not represent the only case in which a subtle idea has been introduced into mathematics by a symbolism which in its origin was dictated by practical convenience.

Thus the first use of 0 was to make the arabic notation possible—no slight service. We can imagine that when it had been introduced for this purpose, practical men, of the sort who dislike fanciful ideas, deprecated the silly habit of identifying it with a number zero. But they were wrong, as such men always are when they desert their proper function of masticating food which others have prepared. For the next service performed by the symbol 0 essentially depends upon assigning to it the function of representing the number zero.

This second symbolic use is at first sight so absurdly simple that it is difficult to make a beginner realize its importance. Let us start with a simple example. In Chapter II. we mentioned the correlation between two variable numbers x and y represented by the equation $x+y=1$. This can be represented in an indefinite number of ways; for example, $x=1-y$, $y=1-x$, $2x+3y-1=x+2y$, and so on. But the important way of stating it is

$$x+y-1=0.$$

Similarly the important way of writing the equation $x=1$ is $x-1=0$, and of representing the equation $3x-2=2x^2$ is $2x^2-3x+2=0$. The point is that all the symbols which represent variables, *e.g.* x and y, and the symbols representing some definite number other than zero, such as 1 or 2 in the examples above, are written on the left-hand side, so that the whole left-hand side is equated to the number zero. The first man to do this is said to have been Thomas Harriot, born at Oxford in 1560 and died in 1621. But what is the importance of this simple symbolic procedure? It made possible the growth of the modern conception of *algebraic form.*

This is an idea to which we shall have continually to recur; it is not going too far to say that no part of modern mathematics can be properly understood without constant recurrence to it. The conception of form is so general that it is difficult to characterize it in abstract terms. At this stage we shall do better merely to consider examples. Thus the equations $2x-3=0$, $x-1=0$, $5x-6=0$, are all equations of the same form, namely, equations involving one unknown x, which is not multiplied by itself, so that x^2, x^3, etc., do not appear. Again $3x^2-2x+1=0$, $x^2-3x+2=0$, $x^2-4=0$, are all equations of the same form, namely, equations involving one unknown x in which $x \times x$, that is x^2, appears. These equations are called quadratic equations. Similarly cubic equations, in which x^3 appears, yield another form, and so on. Among the three quadratic equations given above there is a minor difference between the last equation, $x^2-4=0$, and the preceding two equations, due to the fact that x (as distinct from x^2) does not appear in the last and does in the other two. This distinction is very unimportant in comparison with the great fact that they are all three quadratic equations.

Then further there are the forms of equation stating correlations between two variables; for example, $x+y-1=0$, $2x+3y-8=0$, and so on. These are examples of what is called the *linear* form of equation. The reason for this name of "linear" is that the graphic method of representation, which is explained at the end of Chapter II, always represents such equations by a straight line. Then there are other forms for two variables—for example, the quadratic form, the cubic form, and so on. But the point which we here insist upon is that this study of form is facilitated, and, indeed, made possible, by the standard method of writing equations with the symbol 0 on the right-hand side.

There is yet another function performed by 0 in relation to the study of form. Whatever number x may be, $0 \times x=0$, and $x+0=x$. By means of these properties minor differences of form can be assimilated. Thus the difference mentioned above between the quadratic equations $x^2-3x+2=0$, and $x^2-4=0$, can be obliterated by writing the latter equation in the form $x^2+(0\times x)-4=0$. For, by the laws stated above, $x^2+(0\times x)-4=x^2+0-4=x^2-4$. Hence the equation $x^2-4=0$, is merely representative of a particular class of quadratic equations and belongs to the same general form as does $x^2-3x+2=0$.

For these three reasons the symbol 0, representing the number zero, is essential to modern mathematics. It has rendered possible types of investigation which would have been impossible without it.

The symbolism of mathematics is in truth the outcome of the general ideas which dominate the science. We have now two such general ideas before us, that of the variable and that of algebraic form. The junction of these concepts has imposed on mathematics another type of symbolism almost quaint in its character, but none the less effective. We have seen that an equation involving two variables, x and y, represents a particular correlation between the pair of variables. Thus $x+y-1=0$ represents one definite correlation, and $3x+2y-5=0$ represents another definite correlation between the variables x and y; and both correlations have the form of what we have called linear correlations. But now, how can we represent *any* linear correlation between the variable numbers x and y? Here we want to symbolize *any* linear correlation; just as x symbolizes *any* number. This is done by turning the numbers which occur in the definite correlation $3x+2y-5=0$ into letters. We obtain $ax+by-c=0$. Here a, b, c, stand for variable numbers just as do x and y: but there is a difference in the use of the two sets of variables. We study the general properties of the relationship between x and y while a, b, and c have unchanged values. We do not determine what the values of a, b, and c are; but whatever they are, they remain fixed while we study the relation between the variables x and y for the whole group of possible values of x and y. But when we have obtained the properties of this correlation, we note that, because a, b, and c have not in fact been determined, we have proved properties which must belong to *any* such relation. Thus, by now varying a, b, and c, we arrive at the idea that $ax+by-c=0$ represents a variable linear correlation between x and y. In comparison with x and y, the three variables a, b, and c are called constants. Variables used in this way are sometimes also called parameters.

Now, mathematicians habitually save the trouble of explaining which of their variables are to be treated as "constants," and which as variables, considered as correlated in their equations, by using letters at the end of the alphabet for the "variable" variables, and letters at the beginning of the alphabet for the "constant" variables, or parameters. The two systems meet naturally about the middle of the alphabet. Sometimes a word or two of explanation is necessary; but as a matter of fact custom and common sense are usually sufficient, and surprisingly little confusion is caused by a procedure which seems so lax.

The result of this continual elimination of definite numbers by successive layers of parameters is that the amount of arithmetic performed by mathematicians is extremely small. Many mathematicians dislike all numerical computation and are not particularly expert at it. The territory of arithmetic ends where the two ideas of "variables" and of "algebraic form" commence their sway.

CHAPTER VI

GENERALIZATIONS OF NUMBER

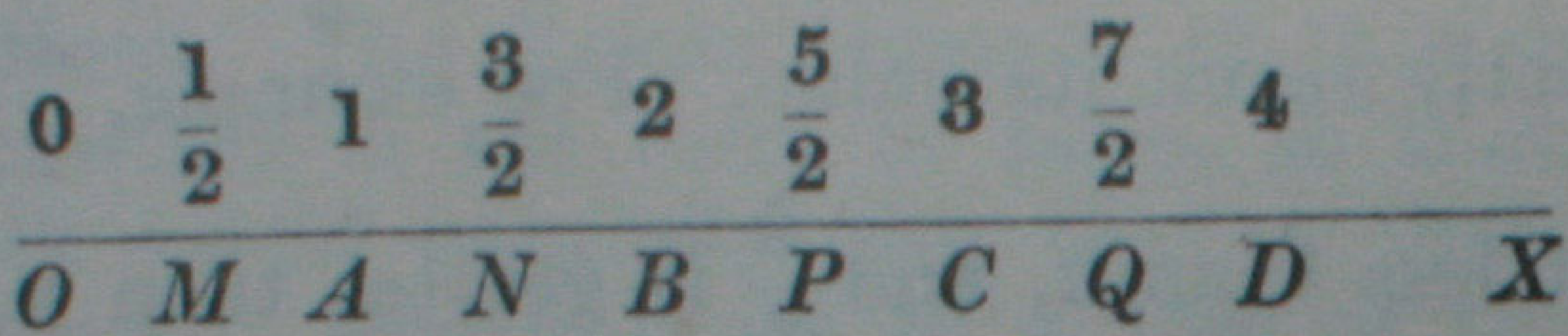
One great peculiarity of mathematics is the set of allied ideas which have been invented in connection with the integral numbers from which we started. These ideas may be called extensions or generalizations of number. In the first place there is the idea of fractions. The earliest treatise on arithmetic which we possess was written by an Egyptian priest, named Ahmes, between 1700 B.C. and 1100 B.C., and it is probably a copy of a much older work. It deals largely with the properties of fractions. It appears, therefore, that this concept was developed very early in the history of mathematics. Indeed the subject is a very obvious one. To divide a field into three equal parts, and to take two of the parts, must be a type of operation which had often occurred. Accordingly, we need not be surprised that the men of remote civilizations were familiar with the idea of two-thirds, and with allied notions. Thus as the first generalization of number we place the concept of fractions. The Greeks thought of this subject rather in the form of ratio, so that a Greek would naturally say that a line of two feet in length bears to a line of three feet in length the ratio of 2 to 3. Under the influence of our algebraic notation we would more often say that one line was two-thirds of the other in length, and would think of two-thirds as a numerical multiplier.

In connection with the theory of ratio, or fractions, the Greeks made a great discovery, which has been the occasion of a large amount of philosophical as well as mathematical thought. They found out the existence of "incommensurable" ratios. They proved, in fact, during the course of their geometrical investigations that, starting with a line of any length, other lines must exist whose lengths do not bear to the original length the ratio of any pair of integers—or, in other words, that lengths exist which are not any exact fraction of the original length.

For example, the diagonal of a square cannot be expressed as any fraction of the side of the same square; in our modern notation the length of the diagonal is $\sqrt{2}$ times the length of the side. But there is no fraction which exactly represents $\sqrt{2}$. We can approximate to $\sqrt{2}$ as closely as we like, but we never exactly reach its value. For example, $\frac{49}{25}$ is just less than 2, and $\frac{9}{4}$ is greater than 2, so that $\sqrt{2}$ lies between $\frac{7}{5}$ and $\frac{3}{2}$. But the best systematic way of approximating to $\sqrt{2}$ in obtaining a series of decimal fractions, each bigger than the last, is by the ordinary method of extracting the square root; thus the series is 1, $\frac{14}{10}$, $\frac{141}{100}$, $\frac{1414}{1000}$, and so on.

Ratios of this sort are called by the Greeks incommensurable. They have excited from the time of the Greeks onwards a great deal of philosophic discussion, and the difficulties connected with them have only recently been cleared up.

We will put the incommensurable ratios with the fractions, and consider the whole set of integral numbers, fractional numbers, and incommensurable numbers as forming one class of numbers which we will call "real numbers." We always think of the real numbers as arranged in order of magnitude, starting from zero and going upwards, and becoming indefinitely larger and larger as we proceed. The real numbers are conveniently represented by points on a line. Let *OX* be



any line bounded at *O* and stretching away indefinitely in the direction *OX*. Take any convenient point, *A*, on it, so that *OA* represents the unit length; and divide off lengths *AB*, *BC*, *CD*, and so on, each equal to *OA*. Then the point *O* represents the number 0, *A* the number 1, *B* the number 2, and so on. In fact the number represented by any point is the measure of its distance from *O*, in terms of the unit length *OA*. The points between *O* and *A* represent the proper fractions and the incommensurable numbers less than 1; the middle point of *OA* represents $\frac{1}{2}$, that of *AB* represents $\frac{3}{2}$, that of *BC* represents $\frac{5}{2}$, and so on. In this way every point on *OX* represents some one real number, and every real number is represented by some one point on *OX*.

The series (or row) of points along *OX*, starting from *O* and moving regularly in the direction from *O* to *X*, represents the real numbers as arranged in an ascending order of size, starting from zero and continually increasing as we go on.

All this seems simple enough, but even at this stage there are some interesting ideas to be got at by dwelling on these obvious facts. Consider the series of points which represent the integral numbers only, namely, the points, *O*, *A*, *B*, *C*, *D*, etc. Here there is a first point *O*, a definite next point, *A*, and each point, such as *A* or *B*, has one definite immediate predecessor and one definite immediate successor, with the exception of *O*, which has no predecessor; also the series goes on indefinitely without end. This sort of order is called the type of order of the integers; its essence is the possession of next-door neighbours on either side with the exception of No. 1 in the row. Again consider the integers and fractions together, omitting the points which correspond to the incommensurable ratios. The sort of serial order which we now obtain is quite different. There is a first term *O*; but no term has any immediate predecessor or immediate successor. This is easily seen to be the case, for between any two fractions we can always find another fraction intermediate in value. One very simple way of doing this is to add the fractions together and to halve the result. For example, between $\frac{2}{3}$ and $\frac{3}{4}$, the fraction $\frac{1}{2}(\frac{2}{3}+\frac{3}{4})$, that is $\frac{17}{24}$, lies; and between $\frac{2}{3}$ and $\frac{17}{24}$ the fraction $\frac{1}{2}(\frac{2}{3}+\frac{17}{24})$, that is $\frac{33}{48}$, lies; and so on indefinitely. Because of this property the series is said to be "compact." There is no end point to the series, which increases indefinitely without limit as we go along the line OX. It would seem at first sight as though the type of series got in this way from the fractions, always including the integers, would be the same as that got from all the real numbers, integers, fractions, and incommensurables taken together, that is, from all the points on the line OX. All that we have hitherto said about the series of fractions applies equally well to the series of all real numbers. But there are important differences which we now proceed to develop. The absence of the incommensurables from the series of fractions leaves an absence of end-points to certain classes. Thus, consider the incommensurable $\sqrt{2}$. In the series of real numbers this stands between all the numbers whose squares are less than 2, and all the numbers whose squares are greater than 2. But keeping to the series of fractions alone and not thinking of the incommensurables, so that we cannot bring in $\sqrt{2}$, there is no fraction which has the property of dividing off the series into two parts in this way, *i.e.* so that all the members on one side have their squares less than 2, and on the other side greater than 2. Hence in the series of fractions there is a quasi-gap where $\sqrt{2}$ ought to come. This presence of quasi-gaps in the series of fractions may seem a small matter; but any mathematician, who happens to read this, knows that the possible absence of limits or maxima to a class of numbers, which yet does not spread over the whole series of numbers, is no small evil. It is to avoid this difficulty that recourse is had to the incommensurables, so as to obtain a complete series with no gaps.

There is another even more fundamental difference between the two series. We can rearrange the fractions in a series like that of the integers, that is, with a first term, and such that each term has an immediate successor and (except the first term) an immediate predecessor. We can show how this can be done. Let every term in the series of fractions and integers be written in the fractional form by writing $\frac{1}{1}$ for 1, $\frac{2}{1}$ for 2, and so on for all the integers, excluding 0. Also for the moment we will reckon fractions which are equal in value but not reduced to their lowest terms as distinct; so that, for example, until further notice $\frac{2}{3}$, $\frac{4}{6}$, $\frac{6}{9}$, $\frac{8}{12}$, etc., are all reckoned as distinct. Now group the fractions into classes by adding together the numerator and denominator of each term. For the sake of brevity call this sum of the numerator and denominator of a fraction its index. Thus 7 is the index of $\frac{4}{3}$, and also of $\frac{3}{4}$, and of $\frac{2}{5}$. Let the fractions in each class be all fractions which have some specified index, which may therefore also be called the class index. Now arrange these classes in the order of magnitude of their indices. The first class has the index 2, and its only member is $\frac{1}{1}$; the second class has the index 3, and its members are $\frac{1}{2}$ and $\frac{2}{1}$; the third class has the index 4, and its members are $\frac{1}{3}$, $\frac{2}{2}$, $\frac{3}{1}$; the fourth class has the index 5, and its members are $\frac{1}{4}$, $\frac{2}{3}$, $\frac{3}{2}$, $\frac{4}{1}$; and so on. It is easy to see that the number of members (still including fractions not in their lowest terms) belonging to any class is one less than its index. Also the members of any one class can be arranged in order by taking the first member to be the fraction with numerator 1, the second member to have the numerator 2, and so on, up to $(n-1)$ where n is the index. Thus for the class of index n, the members appear in the order.

$\frac{1}{n-1}, \frac{2}{n-2}, \frac{3}{n-3}, \ldots, \frac{n-1}{1}$. The members of the first four classes have in fact been mentioned in this order. Thus the whole set of fractions have now been arranged in an order like that of the integers. It runs thus

$$\frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{1}{3}, \left[\frac{2}{2}\right], \frac{3}{1}, \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1}, \ldots,$$

$$\frac{n-2}{1}, \frac{1}{n-1}, \frac{2}{n-2}, \frac{3}{n-3}, \ldots, \frac{n-1}{1}, \frac{1}{n},$$

and so on.

Now we can get rid of all repetitions of fractions of the same value by simply striking them out whenever they appear after their first occurrence. In the few initial terms written down above, $\frac{2}{2}$ which is enclosed above in square brackets is the only fraction not in its lowest terms. It has occurred before as $\frac{1}{1}$. Thus this must be struck out. But the series is still left with the same properties, namely, (*a*) there is a first term, (*b*) each term has next-door neighbours, (*c*) the series goes on without end.

It can be proved that it is not possible to arrange the whole series of real numbers in this way. This curious fact was discovered by Georg Cantor, a German mathematician still living; it is of the utmost importance in the philosophy of mathematical ideas. We are here in fact touching on the fringe of the great problems of the meaning of continuity and of infinity.

Another extension of number comes from the introduction of the idea of what has been variously named an operation or a step, names which are respectively appropriate from slightly different points of view. We will start with a particular case. Consider the statement $2+3=5$. We add 3 to 2 and obtain 5. Think of the operation of adding 3: let this be denoted by $+3$. Again $4-3=1$. Think of the operation of subtracting 3: let this be denoted by -3. Thus instead of considering the real numbers in themselves, we consider the *operations* of adding or subtracting them: instead of $\sqrt{2}$, we consider $+\sqrt{2}$ and $-\sqrt{2}$, namely the operations of adding $\sqrt{2}$ and of subtracting $\sqrt{2}$. Then we can add these operations, of course in a different sense of addition to that in which we add numbers. The sum of two operations is the single operation which has the same effect as the two operations applied successively. In what order are the two operations to be applied? The answer is that it is indifferent, since for example

$$2+3+1=2+1+3;$$

so that the addition of the steps $+3$ and $+1$ is commutative.

Mathematicians have a habit, which is puzzling to those engaged in tracing out meanings, but is very convenient in practice, of using the same symbol in different though allied senses. The one essential requisite for a symbol in their eyes is that, whatever its possible varieties of meaning, the formal laws for its use shall always be the same. In accordance with this habit the addition of operations is denoted by + as well as the addition of numbers. Accordingly we can write

$$(+3)+(+1)=+4;$$

where the middle + on the left-hand side denotes the addition of the operations +3 and +1. But, furthermore, we need not be so very pedantic in our symbolism, except in the rare instances when we are directly tracing meanings; thus we always drop the first + of a line and the brackets, and never write two + signs running. So the above equation becomes

$$3+1=4,$$

which we interpret as simple numerical addition, or as the more elaborate addition of operations which is fully expressed in the previous way of writing the equation, or lastly as expressing the result of applying the operation +1 to the number 3 and obtaining the number 4. Any interpretation which is possible is always correct. But the only interpretation which is always possible, under certain conditions, is that of operations. The other interpretations often give nonsensical results.

This leads us at once to a question, which must have been rising insistently in the reader's mind: What is the use of all this elaboration? At this point our friend, the practical man, will surely step in and insist on sweeping away all these silly cobwebs of the brain. The answer is that what the mathematician is seeking is Generality. This is an idea worthy to be placed beside the notions of the Variable and of Form so far as concerns its importance in governing mathematical procedure. Any limitation whatsoever upon the generality of theorems, or of proofs, or of interpretation is abhorrent to the mathematical instinct. These three notions, of the variable, of form, and of generality, compose a sort of mathematical trinity which preside over the whole subject. They all really spring from the same root, namely from the abstract nature of the science.

Let us see how generality is gained by the introduction of this idea of operations. Take the equation $x+1=3$; the solution is $x=2$. Here we can interpret our symbols as mere numbers, and the recourse to "operations" is entirely unnecessary. But, if x is a mere number, the equation $x+3=1$ is nonsense. For x should be the number of things which remain when you have taken 3 things away from 1 thing; and no such procedure is possible. At this point our idea of algebraic form steps in, itself only generalization under another aspect. We consider, therefore, the