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 * :—

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

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.