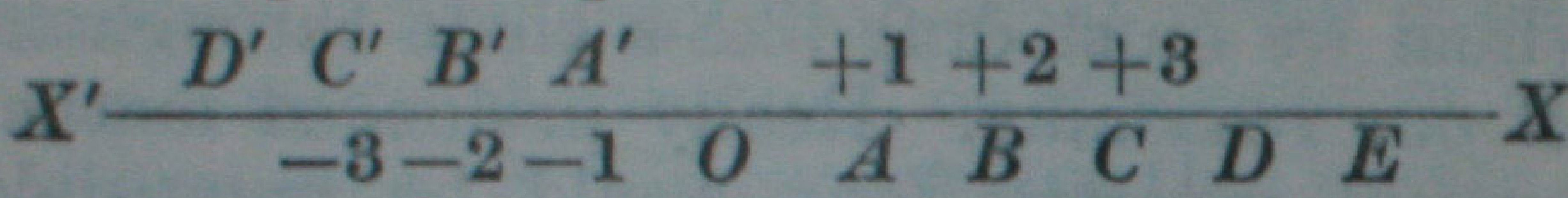
general equation of the same form as $x+1=3$. This equation is $x+a=b$, and its solution is $x=b-a$. Here our difficulties become acute; for this form can only be used for the numerical interpretation so long as b is greater than a, and we cannot say without qualification that a and b may be any constants. In other words we have introduced a limitation on the variability of the "constants" a and b, which we must drag like a chain throughout all our reasoning. Really prolonged mathematical investigations would be impossible under such conditions. Every equation would at last be buried under a pile of limitations. But if we now interpret our symbols as "operations," all limitation vanishes like magic. The equation $x+1=3$ gives $x=+2$, the equation $x+3=1$ gives $x=-2$, the equation $x+a=b$ gives $x=b-a$ which is an operation of addition or subtraction as the case may be. We need never decide whether $b-a$ represents the operation of addition or of subtraction, for the rules of procedure with the symbols are the same in either case.

It does not fall within the plan of this work to write a detailed chapter of elementary algebra. Our object is merely to make plain the fundamental ideas which guide the formation of the science. Accordingly we do not further explain the detailed rules by which the "positive and negative numbers" are multiplied and otherwise combined. We have explained above that positive and negative numbers are operations. They have also been called "steps." Thus +3 is the step by which we go from 2 to 5, and −3 is the step backwards by which we go from 5 to 2. Consider the line OX divided in the way explained in the earlier part of the chapter, so that its points represent numbers. Then +2



is the step from O to B, or from A to C, or (if the divisions are taken backwards along OX') from C' to A', or from D' to B', and so on. Similarly −2 is the step from O to B', or from B' to D', or from B to O, or from C to A.

We may consider the point which is reached by a step from O, as representative of that step. Thus A represents +1, B represents +2, A' represents −1, B' represents −2, and so on. It will be noted that, whereas previously with the mere "unsigned" real numbers the points on one side of O only, namely along OX, were representative of numbers, now with steps every point on the whole line stretching on both sides of O is representative of a step. This is a pictorial representation of the superior generality introduced by the positive and negative numbers, namely the operations or steps. These "signed" numbers are also particular cases of what have been called vectors (from the Latin *veho*, I draw or carry). For we may think of a particle as carried from O to A, or from A to B.

In suggesting a few pages ago that the practical man would object to the subtlety involved by the introduction of the positive and negative numbers, we were libelling that excellent individual. For in truth we are on the scene of one of his greatest triumphs. If the truth must be confessed, it was the practical man himself who first employed the actual symbols + and −. Their origin is not very certain, but it seems most probable that they arose from the marks chalked on chests of goods in German warehouses, to denote excess or defect from some standard weight. The earliest notice of them occurs in a book published at Leipzig, in A.D. 1489. They seem first to have been employed in mathematics by a German mathematician, Stifel, in a book published at Nuremburg in 1544 A.D. But then it is only recently that the Germans have come to be looked on as emphatically a practical nation. There is an old epigram which assigns the empire of the sea to the English, of the land to the French, and of the clouds to the Germans. Surely it was from the clouds that the Germans fetched + and

—; the ideas which these symbols have generated are much too important for the welfare of humanity to have come from the sea or from the land.

The possibilities of application of the positive and negative numbers are very obvious. If lengths in one direction are represented by positive numbers, those in the opposite direction are represented by negative numbers. If a velocity in one direction is positive, that in the opposite direction is negative. If a rotation round a dial in the opposite direction to the hands of a clock (anti-clockwise) is positive, that in the clockwise direction is negative. If a balance at the bank is positive, an overdraft is negative. If vitreous electrification is positive, resinous electrification is negative. Indeed, in this latter case, the terms positive electrification and negative electrification, considered as mere names, have practically driven out the other terms. An endless series of examples could be given. The idea of positive and negative numbers has been practically the most successful of mathematical subtleties.

CHAPTER VII

IMAGINARY NUMBERS

If the mathematical ideas dealt with in the last chapter have been a popular success, those of the present chapter have excited almost as much general attention. But their success has been of a different character, it has been what the French term a *succès de scandale.* Not only the practical man, but also men of letters and philosophers have expressed their bewilderment at the devotion of mathematicians to mysterious entities which by their very name are confessed to be imaginary. At this point it may be useful to observe that a certain type of intellect is always worrying itself and others by discussion as to the applicability of technical terms. Are the incommensurable numbers properly called numbers? Are the positive and negative numbers really numbers? Are the imaginary numbers imaginary, and are they numbers?—are types of such futile questions. Now, it cannot be too clearly understood that, in science, technical terms are names arbitrarily assigned, like Christian names to children. There can be no question of the names being right or wrong. They may be judicious or injudicious; for they can sometimes be so arranged as to be easy to remember, or so as to suggest relevant and important ideas. But the essential principle involved was quite clearly enunciated in Wonderland to Alice by Humpty Dumpty, when he told her, à propos of his use of words, "I pay them extra and make them mean what I like." So we will not bother as to whether imaginary numbers are imaginary, or as to whether they are numbers, but will take the phrase as the arbitrary name of a certain mathematical idea, which we will now endeavour to make plain.

The origin of the conception is in every way similar to that of the positive and negative numbers. In exactly the same way it is due to the three great mathematical ideas of the variable, of algebraic form, and of generalization. The positive and negative numbers arose from the consideration of equations like $x+1=3$, $x+3=1$, and the general form $x+a=b$. Similarly the origin of imaginary numbers is due to equations like $x^2+1=3$, $x^2+3=1$, and $x^2+a=b$. Exactly the same process is gone through. The equation $x^2+1=3$ becomes $x^2=2$, and this has two solutions, either $x=+\sqrt{2}$, or $x=-\sqrt{2}$. The statement that there are these alternative solutions is usually written $x=\pm\sqrt{2}$. So far all is plain sailing, as it was in the previous case. But now an analogous difficulty arises. For the equation $x^2+3=1$ gives $x^2=-2$ and there is no positive or negative number which, when multiplied by itself, will give a negative square. Hence, if our symbols are to mean the ordinary positive or negative numbers, there is no solution to $x^2=-2$, and the equation is in fact nonsense. Thus, finally taking the general form $x^2+a=b$, we find the pair of solutions $x=\pm\sqrt{(b-a)}$, when, and only when, b is not less than a. Accordingly we cannot say unrestrictedly that the "constants" a and b may be any numbers, that is, the "constants" a and b are not, as they ought to be, independent unrestricted "variables"; and so again a host of limitations and restrictions will accumulate round our work as we proceed.

The same task as before therefore awaits us: we must give a new interpretation to our symbols, so that the solutions $\pm\sqrt{(b-a)}$ for the equation $x^2+a=b$ always have meaning. In other words, we require an interpretation of the symbols so that $\sqrt{a}$ always has meaning whether a be positive or negative. Of course, the interpretation must be such that all the ordinary formal laws for addition, subtraction, multiplication, and division hold good; and also it must not interfere with the generality which we have attained by the use of the positive and negative numbers. In fact, it must in a sense include them as special cases. When a is negative we may write $-c^2$ for it, so that c^2 is positive. Then

$$\sqrt{a} = \sqrt{(-c^2)} = \sqrt{\{(-1) \times c^2\}}$$
$$= \sqrt{(-1)}\ \sqrt{c^2} = c\ \sqrt{(-1)}.$$

Hence, if we can so interpret our symbols that $\sqrt{(-1)}$ has a meaning, we have attained our object. Thus $\sqrt{(-1)}$ has come to be looked on as the head and forefront of all the imaginary quantities.

This business of finding an interpretation for $\sqrt{(-1)}$ is a much tougher job than the analogous one of interpreting -1. In fact, while the easier problem was solved almost instinctively as soon as it arose, it at first hardly occurred, even to the greatest mathematicians, that here a problem existed which was perhaps capable of solution. Equations like $x^2 = -3$, when they arose, were simply ruled aside as nonsense.

However, it came to be gradually perceived during the eighteenth century, and even earlier, how very convenient it would be if an interpretation could be assigned to these nonsensical symbols. Formal reasoning with these symbols was gone through, merely assuming that they obeyed the ordinary algebraic laws of transformation; and it was seen that a whole world of interesting results could be attained, if only these symbols might legitimately be used. Many mathematicians were not then very clear as to the logic of their procedure, and an idea gained ground that, in some mysterious way, symbols which mean nothing can by appropriate manipulation yield valid proofs of propositions. Nothing can be more mistaken. A symbol which has not been properly defined is not a symbol at all. It is merely a blot of ink on paper which has an easily recognized shape. Nothing can be proved by a succession of blots, except the existence of a bad pen or a careless writer. It was during this epoch that the epithet "imaginary" came to be applied to $\sqrt{(-1)}$. What these mathematicians had really succeeded in proving were a series of hypothetical propositions, of which this is the blank form: If interpretations exist for $\sqrt{(-1)}$ and for the addition, subtraction, multiplication, and division of $\sqrt{(-1)}$ which make the ordinary algebraic rules (*e.g.* $x+y=y+x$, etc.) to be satisfied, then such and such results follows. It was natural that the mathematicians should not always appreciate the big "If," which ought to have preceded the statements of their results.

As may be expected the interpretation,

when found, was a much more elaborate affair than that of the negative numbers and the reader's attention must be asked for some careful preliminary explanation. We have already come across the representation of a point by two numbers. By the aid of the

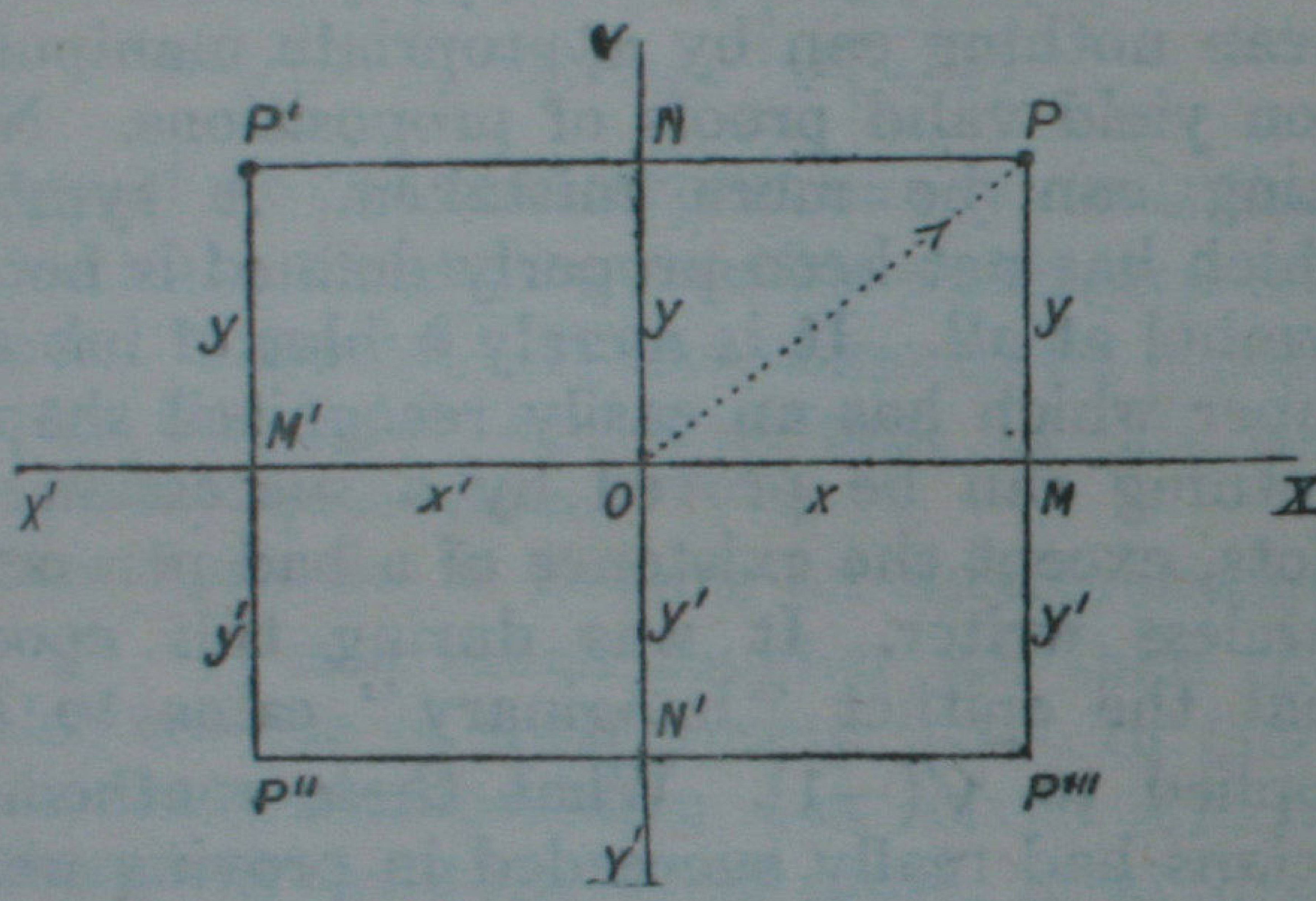


Fig. 8.

positive and negative numbers we can now represent the position of any point in a plane by a pair of such numbers. Thus we take the pair of straight lines XOX' and YOY', at right angles, as the "axes" from which we start all our measurements. Lengths measured along OX and OY are positive, and measured backwards along OX' and OY' are negative. Suppose that a pair of numbers, written in order, *e.g.* $(+3, +1)$, so that there is a first number (+3 in the above example), and a second number (+1 in the above example), represents measurements from O along XOX' for the first number, and along YOY' for the second number. Thus (cf. fig. 9) in (+3, +1) a length of 3 units is to be measured along XOX' in the positive direction, that is from O towards X, and a length +1 measured along YOY' in the positive direction, that is from O towards Y. Similarly in (−3, +1) the length of 3 units is to be measured from O towards X', and of 1 unit from O towards Y. Also in (−3, −1) the two lengths are to be measured along OX' and OY' respectively, and in (+3, −1) along OX and OY' respectively. Let us for the moment call such a pair of numbers an "ordered couple." Then, from the two numbers 1 and 3, eight ordered couples can be generated, namely

$$(+1, +3), (-1, +3), (-1, -3), (+1, -3),$$
$$(+3, +1), (-3, +1), (-3, -1), (+3, -1).$$

Each of these eight "ordered couples" directs a process of measurement along XOX' and YOY' which is different from that directed by any of the others.

The processes of measurement represented by the last four ordered couples, mentioned above, are given pictorially in the figure. The lengths OM and ON together correspond to $(+3, +1)$, the lengths OM' and ON together correspond to $(-3, +1)$, OM' and ON' together to $(-3, -1)$, and OM and ON' together to $(+3, -1)$. But by completing the various rectangles, it is easy to see that the point P completely determines and is determined by the ordered couple

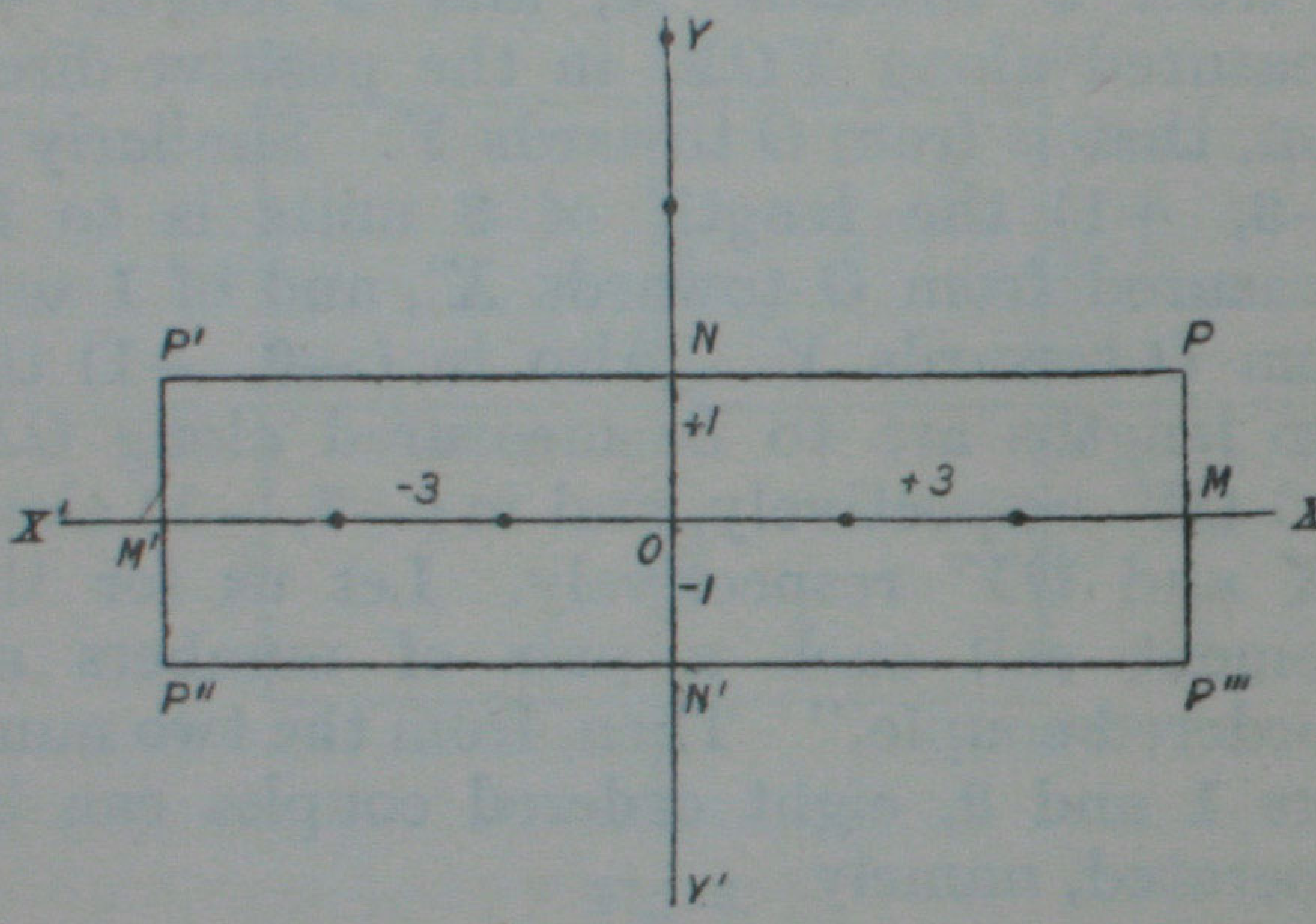


Fig. 9.

$(+3, +1)$, the point P' by $(-3, +1)$, the point P'' by $(-3, -1)$, and the point P''' by $(+3, -1)$. More generally in the previous figure (8), the point P corresponds to the ordered couple (x, y), where x and y in the figure are both assumed to be positive, the point P' corresponds to (x', y), where x' in the figure is assumed to be negative, P'' to (x', y'), and P''' to (x, y'). Thus an ordered couple (x, y), where x and y are any positive or negative numbers, and the corresponding point reciprocally determine each other. It is convenient to introduce some names at this juncture. In the ordered couple (x, y) the first number x is called the "abscissa" of the corresponding point, and the second number y is called the "ordinate" of the point, and the two numbers together are called the "co-ordinates" of the point. The idea of determining the position of a point by its "co-ordinates" was by no means new when the theory of "imaginaries" was being formed. It was due to Descartes, the great French mathematician and philosopher, and appears in his *Discours* published at Leyden in 1637 A.D. The idea of the ordered couple as a thing on its own account is of later growth and is the outcome of the efforts to interpret imaginaries in the most abstract way possible.

It may be noticed as a further illustration of this idea of the ordered couple, that the point M in fig. 9 is the couple $(+3, 0)$, the point N is the couple $(0, +1)$, the point M' the couple $(-3, 0)$, the point N' the couple $(0, -1)$, the point O the couple $(0, 0)$.

Another way of representing the ordered couple (x, y) is to think of it as representing the dotted line OP (cf. fig. 8), rather than the point P. Thus the ordered couple represents a line drawn from an "origin," O, of a certain length and in a certain direction. The line OP may be called the vector line from O to P, or the step from O to P. We see, therefore, that we have in this chapter only extended the interpretation which we gave formerly of the positive and negative numbers. This method of representation by vectors is very useful when we consider the meaning to be assigned to the operations of the addition and multiplication of ordered couples.

We will now go on to this question, and ask what meaning we shall find it convenient to assign to the addition of the two ordered couples (x, y) and (x', y'). The interpretation must, (a) make the result of addition to be another ordered couple, (b) make the operation commutative so that $(x, y)+(x', y')=(x', y')+(x, y)$, ($c$) make the operation associative so that

$$\{(x, y)+(x', y')\} + (u, v) \\ =(x, y)+\{(x', y')+(u, v)\},$$

(d) make the result of subtraction unique, so that when we seek to determine the unknown ordered couple (x, y) so as to satisfy the equation

$$(x, y)+(a, b)=(c, d),$$

there is one and only one answer which we can represent by

$$(x, y)=(c, d)-(a, b).$$

All these requisites are satisfied by taking $(x, y)+(x', y')$ to mean the ordered couple $(x+x', y+y')$. Accordingly by definition we put

$$(x, y)+(x', y')=(x+x', y+y').$$

Notice that here we have adopted the mathematical habit of using the same symbol + in different senses. The + on the left-hand side of the equation has the new meaning of + which we are just defining; while the two +'s on the right-hand side have the meaning of the addition of positive and negative numbers (operations) which was defined in the last chapter. No practical confusion arises from this double use.

As examples of addition we have

$$(+3, +1)+(+2, +6)=(+5, +7),$$
$$(+3, -1)+(-2, -6)=(+1, -7),$$
$$(+3, +1)+(-3, -1)=(0, 0).$$

The meaning of subtraction is now settled for us. We find that

$$(x, y)-(u, v)=(x-u, y-v).$$

Thus

$$(+3, +2)-(+1, +1)=(+2, +1),$$

and

$$(+1, -2)-(+2, -4)=(-1, +2),$$

and

$$(-1, -2)-(+2, +3)=(-3, -5).$$

It is easy to see that

$$(x, y)-(u, v)=(x, y)+(-u, -v).$$

Also

$$(x, y)-(x, y)=(0, 0).$$

Hence (0, 0) is to be looked on as the zero ordered couple. For example

$$(x, y)+(0, 0)=(x, y).$$

The pictorial representation of the addition of ordered couples is surprisingly easy.

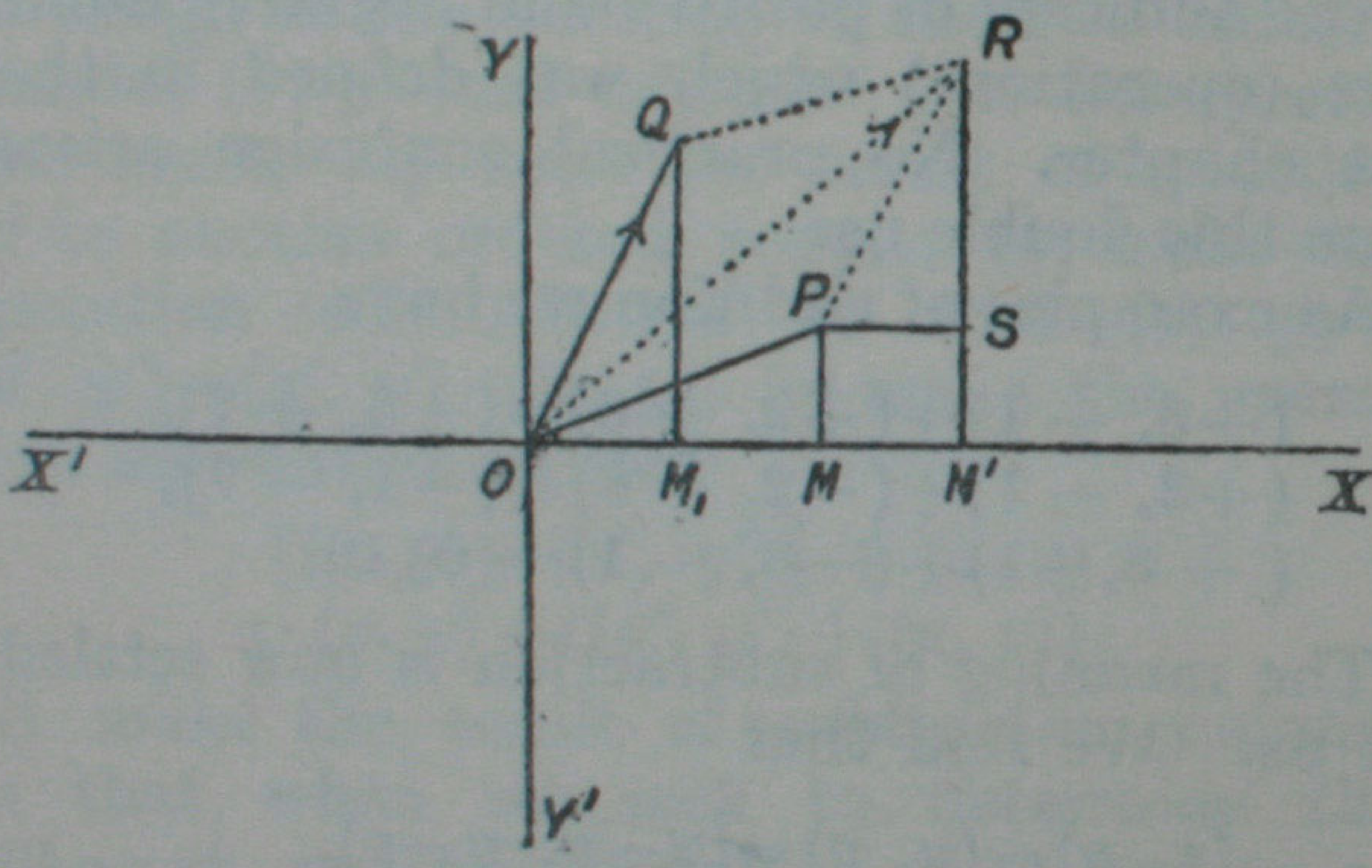


Fig. 10.

Let OP represent (x, y) so that $OM=x$ and $PM=y$; let OQ represent (x_1, y_1) so that $OM_1=x_1$ and $QM_1=y_1$. Complete the parallelogram $OPRQ$ by the dotted lines PR and QR, then the diagonal OR is the ordered couple $(x+x_1, y+y_1)$. For draw PS parallel to OX; then evidently the triangles OQM_1 and PRS are in all respects equal. Hence $MM' = PS = x_1$, and $RS = QM_1$; and therefore

$$OM' = OM + MM' = x + x_1,$$
$$RM' = SM' + RS = y + y_1.$$

Thus OR represents the ordered couple as required. This figure can also be drawn with OP and OQ in other quadrants.

It is at once obvious that we have here come back to the parallelogram law, which was mentioned in Chapter VI., on the laws of motion, as applying to velocities and forces. It will be remembered that, if OP and OQ represent two velocities, a particle is said to be moving with a velocity equal to the two velocities added together if it be moving with the velocity OR. In other words OR is said to be the resultant of the two velocities OP and OQ. Again forces acting at a point of a body can be represented by lines just as velocities can be; and the same parallelogram law holds, namely, that the resultant of the two forces OP and OQ is the force represented by the diagonal OR. It follows that we can look on an ordered couple as representing a velocity or a force, and the rule which we have just given for the addition of ordered couples then represents the fundamental laws of mechanics for the addition of forces and velocities. One of the most fascinating characteristics of mathematics is the surprising way in which the ideas and results of different parts of the subject dovetail into each other. During the discussions of this and the previous chapter we have been guided merely by the most abstract of pure mathematical considerations; and yet at the end of them we have been led back to the most fundamental of all the laws of nature, laws which have to be in the mind of every engineer as he designs an engine, and of every naval architect as he calculates the stability of a ship. It is no paradox to say that in our most theoretical moods we may be nearest to our most practical applications.

CHAPTER VIII

IMAGINARY NUMBERS (*Continued*)

THE definition of the multiplication of ordered couples is guided by exactly the same considerations as is that of their addition. The interpretation of multiplication must be such that

(α) the result is another ordered couple,

(β) the operation is commutative, so that

$$(x, y) \times (x', y') = (x', y') \times (x, y),$$

(γ) the operation is associative, so that

$$\{(x, y) \times (x', y')\} \times (u, v)$$
$$= (x, y) \times \{(x', y') \times (u, v)\},$$

(δ) must make the result of division unique [with an exception for the case of the zero couple (0, 0)], so that when we seek to determine the unknown couple (x, y) so as to satisfy the equation

$$(x, y) \times (a, b) = (c, d),$$

there is one and only one answer, which we can represent by

$$(x, y) = (c, d) \div (a, b), \text{ or by } (x, y) = \frac{(c, d)}{(a, b)}$$

(ϵ) Furthermore the law involving both addition and multiplication, called the distributive law, must be satisfied, namely

$$(x, y) \times \{(a, b) + (c, d)\}$$
$$= \{(x, y) \times (a, b)\} + \{(x, y) \times (c, d)\}.$$

All these conditions (α), (β), (γ), (δ), (ϵ) can be satisfied by an interpretation which, though it looks complicated at first, is capable of a simple geometrical interpretation.

By definition we put

$$(x, y) \times (x', y') = \{(xx' - yy'), (xy' + x'y)\} \quad \text{(A)}$$

This is the definition of the meaning of the symbol $\times$ when it is written between two ordered couples. It follows evidently from this definition that the result of multiplication is another ordered couple, and that the value of the right-hand side of equation (A) is not altered by simultaneously interchanging x with x', and y with y'. Hence conditions (α) and (β) are evidently satisfied. The proof of the satisfaction of (γ), (δ), (ϵ) is equally easy when we have given the geometrical interpretation, which we will proceed to do in a moment. But before doing this it will be interesting to pause and see whether we have attained the object for which all this elaboration was initiated.

We came across equations of the form $x^2 = -3$, to which no solutions could be assigned in terms of positive and negative real numbers. We then found that all our difficulties would vanish if we could interpret the equation $x^2 = -1$, *i.e.*, if we could so define $\sqrt{(-1)}$ that $\sqrt{(-1)} \times \sqrt{(-1)} = -1$.

Now let us consider the three special ordered couples * (0,0), (1,0), and (0,1).

We have already proved that

$$(x, y)+(0, 0)=(x, y).$$

Furthermore we now have

$$(x, y) \times (0, 0)=(0, 0).$$

Hence both for addition and for multiplication the couple (0,0) plays the part of zero in elementary arithmetic and algebra; compare the above equations with $x+0=x$, and $x \times 0=0$.

Again consider (1, 0): this plays the part of 1 in elementary arithmetic and algebra. In these elementary sciences the special characteristic of 1 is that $x \times 1=x$, for all values of x. Now by our law of multiplication

$$(x, y) \times (1, 0)= \{(x-0), (y+0)\} = (x, y).$$

Thus (1, 0) is the unit couple.

* For the future we follow the custom of omitting the + sign wherever possible, thus (1,0) stands for (+1,0) and (0,1) for (0,+1).

Finally consider (0,1): this will interpret for us the symbol $\sqrt{(-1)}$. The symbol must therefore possess the characteristic property that $\sqrt{(-1)} \times \sqrt{(-1)} = -1$. Now by the law of multiplication for ordered couples

$$(0,1) \times (0,1) = \{(0-1), (0+0)\} = (-1, 0).$$

But (1,0) is the unit couple, and (−1, 0) is the negative unit couple; so that (0,1) has the desired property. There are, however, two roots of −1 to be provided for, namely $\pm\sqrt{(-1)}$. Consider (0, −1); here again remembering that $(-1)^2 = 1$, we find, $(0, -1) \times (0, -1) = (-1, 0)$.

Thus (0, −1) is the other square root of $\sqrt{(-1)}$. Accordingly the ordered couples (0,1) and (0,−1) are the interpretations of $\pm\sqrt{(-1)}$ in terms of ordered couples. But which corresponds to which? Does (0,1) correspond to $+\sqrt{(-1)}$ and (0, −1) to $-\sqrt{(-1)}$, or (0,1) to $-\sqrt{(-1)}$, and (0, −1) to $+\sqrt{(-1)}$? The answer is that it is perfectly indifferent which symbolism we adopt.

The ordered couples can be divided into three types, (i) the "complex imaginary" type (x,y), in which neither x nor y is zero; (ii) the "real" type $(x,0)$; (iii) the "pure imaginary" type $(0,y)$. Let us consider the relations of these types to each other. First multiply together the "complex imaginary" couple (x,y) and the "real" couple $(a,0)$, we find

$$(a,0)\times(x,y)=(ax,\ ay).$$

Thus the effect is merely to multiply each term of the couple (x,y) by the positive or negative real number a.

Secondly, multiply together the "complex imaginary" couple (x,y) and the "pure imaginary" couple $(0,b)$, we find

$$(0,b)\times(x,y)=(-by,\ bx).$$

Here the effect is more complicated, and is best comprehended in the geometrical interpretation to which we proceed after noting three yet more special cases.

Thirdly, we multiply the "real" couple $(a,0)$ by the imaginary $(0,b)$ and obtain

$$(a,0)\times(0,b)=(0,ab).$$

Fourthly, we multiply the two "real" couples $(a,0)$ and $(a',0)$ and obtain

$$(a,0)\times(a',0)=(aa',0).$$

Fifthly, we multiply the two "imaginary couples" $(0,b)$ and $(0,b)$ and obtain

$$(0,b)\times(0,b')=(-bb',\ 0).$$

We now turn to the geometrical interpretation, beginning first with some special cases.

Take the couples (1,3) and (2,0) and consider the equation

$$(2,0) \times (1,3) = (2,6)$$

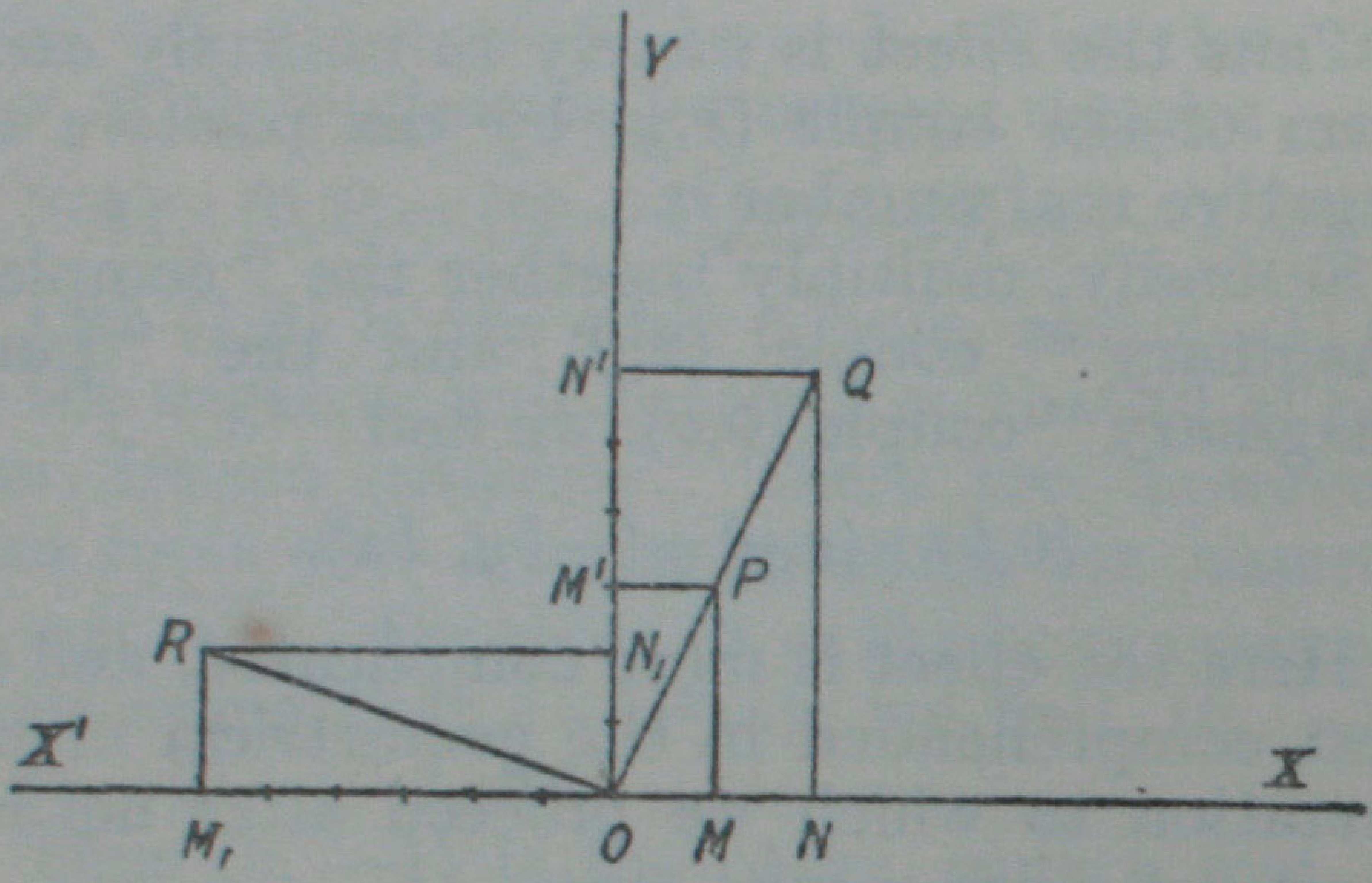


Fig. 11.

In the diagram (fig. 11) the vector *OP* represents (1, 3), and the vector *ON* represents (2,0), and the vector *OQ* represents (2,6). Thus the product $(2,0) \times (1,3)$ is found geometrically by taking the length of the vector *OQ* to be the product of the lengths of the vectors *OP* and *ON*, and (in this case) by producing *OP* to *Q* to be of the required length. Again, consider the product $(0,2) \times (1,3)$, we have

$$(0, 2) \times (1, 3) = (-6, 2)$$

The vector ON_1, corresponds to (0, 2) and the vector *OR* to $(-6,2)$. Thus *OR* which represents the new product is at right angles to OQ and of the same length. Notice that we have the same law regulating the length of OQ as in the previous case, namely, that its length is the product of the lengths of the two vectors which are multiplied together; but now that we have ON_1 along the "ordinate" axis OY, instead of ON along the "abscissa" axis OX, the direction of OP has been turned through a right-angle.

Hitherto in these examples of multiplication we have looked on the vector OP as modified by the vectors ON and ON_1. We shall get a clue to the general law for the direction by inverting the way of thought, and by thinking of the vectors ON and ON_1 as modified by the vector OP. The law for the length remains unaffected; the resultant length is the length of the product of the two vectors. The new direction for the enlarged ON (*i.e.* OQ) is found by rotating it in the (anti-clockwise) direction of rotation from OX towards OY through an angle equal to the angle XOP: it is an accident of this particular case that this rotation makes OQ lie along the line OP. Again consider the product of ON_1 and OP; the new direction for the enlarged ON_1 (*i.e.* OR) is found by rotating ON in the anti-clockwise direction of rotation through an angle equal to the angle XOP, namely, the angle N_1OR is equal to the angle XOP.

The general rule for the geometrical representation of multiplication can now be enunciated thus:

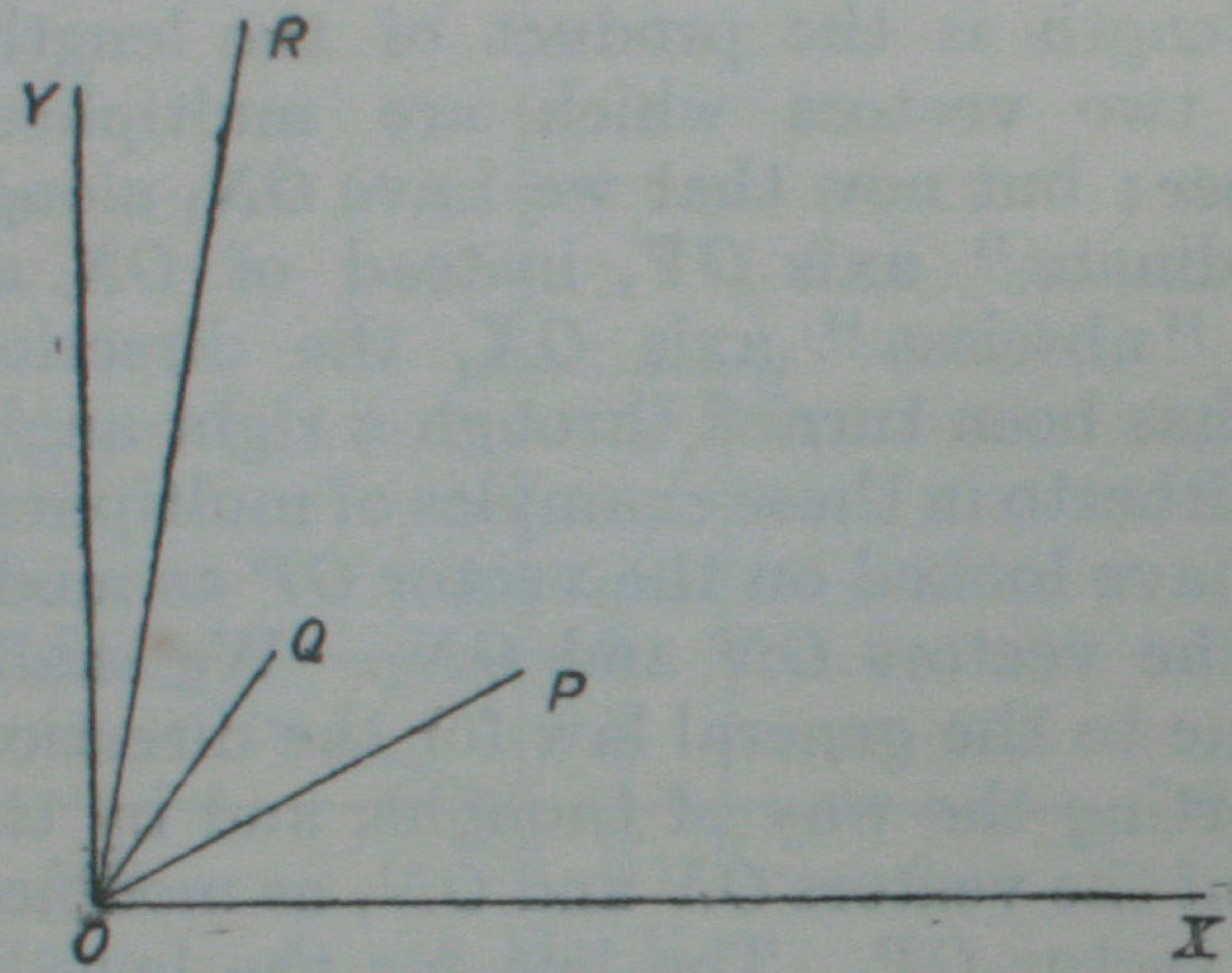


Fig. 12.

The product of the two vectors OP and OQ is a vector OR, whose length is the product of the lengths of OP and OQ and whose direction OR is such that the angle XOR is equal to the sum of the angles XOP and XOQ.

Hence we can conceive the vector OP as making the vector OQ rotate through an angle XOP (*i.e.* the angle $QOR=$ the angle XOP), or the vector OQ as making the vector OP rotate through the angle XOQ (*i.e.* the angle $POR=$the angle XOQ).

We do not prove this general law, as we

should thereby be led into more technical processes of mathematics than falls within the design of this book. But now we can immediately see that the associative law [numbered (γ) above] for multiplication is satisfied. Consider first the length of the resultant vector; this is got by the ordinary process of multiplication for real numbers; and thus the associative law holds for it.

Again, the direction of the resultant vector is got by the mere addition of angles, and the associative law holds for this process also.

So much for multiplication. We have now rapidly indicated, by considering addition and multiplication, how an algebra or "calculus" of vectors in one plane can be constructed, which is such that any two vectors in the plane can be added, or subtracted, and can be multiplied, or divided one by the other.

We have not considered the technical details of all these processes because it would lead us too far into mathematical details; but we have shown the general mode of procedure. When we are interpreting our algebraic symbols in this way, we are said to be employing "imaginary quantities" or "complex quantities." These terms are mere details, and we have far too much to think about to stop to enquire whether they are or are not very happily chosen.

The nett result of our investigations is that any equations like $x+3=2$ or $(x+3)^2= -2$ can now always be interpreted into terms of vectors, and solutions found for them. In seeking for such interpretations it is well to note that 3 becomes (3,0), and −2 becomes (−2,0), and x becomes the "unknown" couple (u, v): so the two equations become respectively $(u, v)+(3,0)=(2,0)$, and $\{(u,v)+(3,0)\}^2 = (-2,0)$.

We have now completely solved the initial difficulties which caught our eye as soon as we considered even the elements of algebra. The science as it emerges from the solution is much more complex in ideas than that with which we started. We have, in fact, created a new and entirely different science, which will serve all the purposes for which the old science was invented and many more in addition. But, before we can congratulate ourselves on this result to our labours, we must allay a suspicion which ought by this time to have arisen in the mind of the student. The question which the reader ought to be asking himself is: Where is all this invention of new interpretations going to end? It is true that we have succeeded in interpreting algebra so as always to be able to solve a quadratic equation like $x^2-2x+4=0$; but there are an endless number of other equations, for example, $x^3-2x+4=0$, $x^4+x^3+2=0$, and so on without limit. Have we got to make a new science whenever a new equation appears?

Now, if this were the case, the whole of our preceding investigations, though to some minds they might be amusing, would in truth be of very trifling importance. But the great fact, which has made modern analysis possible, is that, by the aid of this calculus of vectors, every formula which arises can receive its proper interpretation; and the "unknown" quantity in every equation can be shown to indicate some vector. Thus the science is now complete in itself as far as its fundamental ideas are concerned. It was receiving its final form about the same time as when the steam engine was being perfected, and will remain a great and powerful weapon for the achievement of the victory of thought over things when curious specimens of that machine repose in museums in company with the helmets and breastplates of a slightly earlier epoch.

CHAPTER IX

COORDINATE GEOMETRY

The methods and ideas of coordinate geometry have already been employed in the previous chapters. It is now time for us to consider them more closely for their own sake; and in doing so we shall strengthen our hold on other ideas to which we have attained. In the present and succeeding chapters we will go back to the idea of the positive and negative real numbers and will ignore the imaginaries which were introduced in the last two chapters.

We have been perpetually using the idea that, by taking two axes, XOX' and YOY', in a plane, any point P in that plane can be determined in position by a pair of positive or negative numbers x and y, where (cf. fig. 13) x is the length OM and y is the length PM. This conception, simple as it looks, is the main idea of the great subject of coordinate geometry. Its discovery marks a momentous epoch in the history of mathematical thought. It is due (as has been already said) to the philosopher Descartes, and occurred to him as an important mathematical method one morning as he lay in bed. Philosophers, when they have possessed a thorough knowledge of mathematics, have been among those who have enriched the

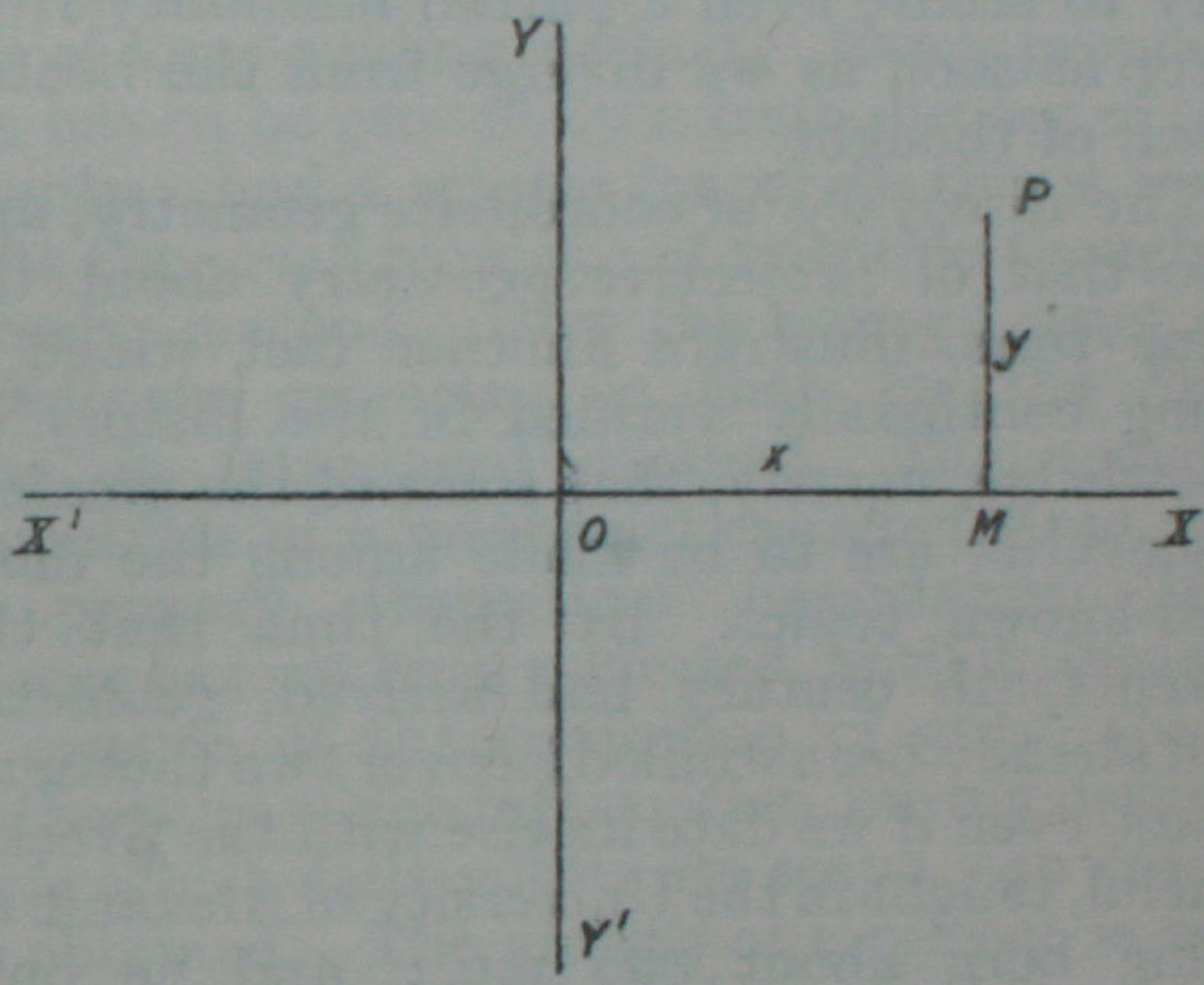


Fig. 13.

science with some of its best ideas. On the other hand it must be said that, with hardly an exception, all the remarks on mathematics made by those philosophers who have possessed but a slight or hasty and late-acquired knowledge of it are entirely worthless, being either trivial or wrong. The fact is a curious one; since the ultimate ideas of mathematics seem, after all, to be very simple, almost childishly so, and to lie well within the province of philosophical thought. Probably their very simplicity is the cause of error; we are not used to think about such simple abstract things, and a long training is necessary to secure even a partial immunity from error as soon as we diverge from the beaten track of thought.

The discovery of coordinate geometry, and also that of projective geometry about the same time, illustrate another fact which is being continually verified in the history of knowledge, namely, that some of the greatest discoveries are to be made among the most well-known topics. By the time that the seventeenth century had arrived, geometry had already been studied for over two thousand years, even if we date its rise with the Greeks. Euclid, taught in the University of Alexandria, being born about 330 B.C.; and he only systematized and extended the work of a long series of predecessors, some of them men of genius. After him generation after generation of mathematicians laboured at the improvement of the subject. Nor did the subject suffer from that fatal bar to progress, namely, that its study was confined to a narrow group of men of similar origin and outlook—quite the contrary was the case; by the seventeenth century it had passed through the minds of Egyptians and Greeks, of Arabs and of Germans. And yet, after all this labour devoted to it through so many ages by such diverse minds its most important secrets were yet to be discovered.

No one can have studied even the elements of elementary geometry without feeling the lack of some guiding method. Every proposition has to be proved by a fresh display of ingenuity; and a science for which this is true lacks the great requisite of scientific thought, namely, method. Now the especial point of coordinate geometry is that for the first time it introduced method. The remote deductions of a mathematical science are not of primary theoretical importance. The science has not been perfected, until it consists in essence of the exhibition of great allied methods by which information, on any desired topic which falls within its scope, can easily be obtained. The growth of a science is not primarily in bulk, but in ideas; and the more the ideas grow, the fewer are the deductions which it is worth while to write down. Unfortunately, mathematics is always encumbered by the repetition in text-books of numberless subsidiary propositions, whose importance has been lost by their absorption into the role of particular cases of more general truths—and, as we have already insisted, generality is the soul of mathematics.

Again, coordinate geometry illustrates another feature of mathematics which has already been pointed out, namely, that mathematical sciences as they develop dovetail into each other, and share the same ideas in common. It is not too much to say that the various branches of mathematics undergo a perpetual process of generalization, and that as they become generalized, they coalesce. Here again the reason springs from the very nature of the science, its generality, that is to say, from the fact that the science deals with the general truths which apply to all things in virtue of their very existence as things. In this connection the interest of coordinate geometry lies in the fact that it relates together geometry, which started as the science of space, and algebra, which has its origin in the science of number.

Let us now recall the main ideas of the two sciences, and then see how they are related by Descartes' method of coordinates. Take algebra in the first place. We will not trouble ourselves about the imaginaries and will think merely of the real numbers with positive or negative signs. The fundamental idea is that of any number, the variable number, which is denoted by a letter and not by any definite numeral. We then proceed to the consideration of correlations between variables. For example, if x and y are two variables, we may conceive them as correlated by the equations $x+y=1$, or by $x-y=1$, or in any one of an indefinite number of other ways. This at once leads to the application of the idea of algebraic form. We think, in fact, of any correlation of some interesting type, thus rising from the initial conception of variable numbers to the secondary conception of variable correlations of numbers. Thus we generalize the correlation $x+y=1$, into the correlation $ax+by=c$. Here a and b and c, being letters, stand for any numbers and are in fact themselves variables. But they are the variables which determine the variable correlation; and the correlation, when determined, correlates the variable numbers x and y. Variables, like a, b, and c above, which are used to determine the correlation are called "constants," or parameters. The use of the term "constant" in this connection for what is really a variable may seem at first sight to be odd; but it is really very natural. For the mathematical investigation is concerned with the relation between the variables x and y, after a, b, c are supposed to have been determined. So in a sense, relatively to x and y, the "constants" a, b, and c are constants. Thus $ax+by=c$ stands for the general example of a certain algebraic form, that is, for a variable correlation belonging to a certain class.

Again we generalize $x^2+y^2=1$ into $ax^2+by^2=c$, or still further into $ax^2+2hxy+by^2=c$, or, still further, into $ax^2+2hxy+by^2+2gx+2fy=c$.
Here again we are led to variable correlations which are indicated by their various algebraic forms.

Now let us turn to geometry. The name of the science at once recalls to our minds the thought of figures and diagrams exhibiting triangles and rectangles and squares and circles, all in special relations to each other. The study of the simple properties of these figures is the subject matter of elementary geometry, as it is rightly presented to the beginner. Yet a moment's thought will show that this is not the true conception of the subject. It may be right for a child to commence his geometrical reasoning on shapes, like triangles and squares, which he has cut out with scissors. What, however, is a triangle? It is a figure marked out and bounded by three bits of three straight lines.

Now the boundary of spaces by bits of lines is a very complicated idea, and not at all one which gives any hope of exhibiting the simple general conceptions which should form the bones of the subject. We want something more simple and more general. It is this obsession with the wrong initial ideas —very natural and good ideas for the creation of first thoughts on the subject—which was the cause of the comparative sterility of the study of the science during so many centuries. Coordinate geometry, and Descartes its inventor, must have the credit of disclosing the true simple objects for geometrical thought.

In the place of a bit of a straight line, let us think of the whole of a straight line throughout its unending length in both directions. This is the sort of general idea from which to start our geometrical investigations. The Greeks never seem to have found any use for this conception which is now fundamental in all modern geometrical thought. Euclid always contemplates a straight line as drawn between two definite points, and is very careful to mention when it is to be produced beyond this segment. He never thinks of the line as an entity given once for all as a whole. This careful definition and limitation, so as to exclude an infinity not immediately apparent to the senses, was very characteristic of the Greeks in all their many activities. It is enshrined in the difference between Greek architecture and Gothic architecture, and between the Greek religion and the modern religion. The spire on a Gothic cathedral and the importance of the unbounded straight line in modern geometry are both emblematic of the transformation of the modern world.

The straight line, considered as a whole, is accordingly the root idea from which modern geometry starts. But then other sorts of lines occur to us, and we arrive at the conception of the complete curve which at every point of it exhibits some uniform characteristic, just as the straight line exhibits at all points the characteristic of straightness. For example, there is the circle which at all points exhibits the characteristic of being at a given distance from its centre, and again there is the ellipse, which is an oval curve, such that the sum of the two distances of any point on it from two fixed points, called its *foci*, is constant for all points on the curve. It is evident that a circle is merely a particular case of an ellipse when the two foci are superposed in the same point; for then the sum of the two distances is merely twice the radius of the circle. The ancients knew the properties of the ellipse and the circle and, of course, considered them as wholes. For example, Euclid never starts with mere segments (*i.e.*, bits) of circles, which are then prolonged. He always considers the whole circle as described. It is unfortunate that the circle is not the true fundamental line in geometry, so that his defective consideration of the straight line might have been of less consequence.

This general idea of a curve which at any point of it exhibits some uniform property is expressed in geometry by the term "locus." A locus is the curve (or surface, if we do not confine ourselves to a plane) formed by points, all of which possess some given property. To every property in relation to each other which points can have, there corresponds some locus, which consists of all the points possessing the property. In investigating the properties of a locus considered as a whole, we consider *any* point or points on the locus. Thus in geometry we again meet with the fundamental idea of the variable. Furthermore, in classifying loci under such headings as straight lines, circles, ellipses, etc., we again find the idea of form.

Accordingly, as in algebra we are concerned with variable numbers, correlations between variable numbers, and the classification of correlations into types by the idea of algebraic form; so in geometry we are concerned with variable points, variable points satisfying some condition so as form to a locus, and the classification of *loci* into types by the idea of conditions of the same form.

Now, the essence of coordinate geometry is the identification of the algebraic correlation with the geometrical locus. The point on a plane is represented in algebra by its two coordinates, x and y, and the condition satisfied by any point on the locus is represented by the corresponding correlation between x and y. Finally to correlations expressible in some general algebraic form, such as $ax+by=c$, there correspond loci of some general type, whose geometrical conditions are all of the same form. We have thus arrived at a position where we can effect a complete interchange in ideas and results between the two sciences. Each science throws light on the other, and itself gains immeasurably in power. It is impossible not to feel stirred at the thought of the emotions of men at certain historic moments of adventure and discovery—Columbus when he first saw the Western shore, Pizarro when he stared at the Pacific Ocean, Franklin when the electric spark came from the string of his kite, Galileo when he first turned his telescope to the heavens. Such moments are also granted to students in the abstract regions of thought, and high among them must be placed the morning when Descartes lay in bed and invented the method of coordinate geometry.

When one has once grasped the idea of coordinate geometry, the immediate question which starts to the mind is, What sort of loci correspond to the well-known algebraic forms? For example, the simplest among the general types of algebraic forms is $ax+by=c$. The sort of locus which corresponds to this is a straight line, and conversely to every straight line there corresponds an equation of this form. It is fortunate that the simplest among the geometrical loci should correspond to the simplest among the algebraic forms. Indeed, it is this general correspondence of geometrical and algebraic simplicity which gives to the whole subject its power. It springs from the fact that the connection between geometry and algebra is not casual and artificial, but deep-seated and essential. The equation which corresponds to a locus is called the equation "of" (or "to") the locus. Some examples of equations of straight lines will illustrate the subject.

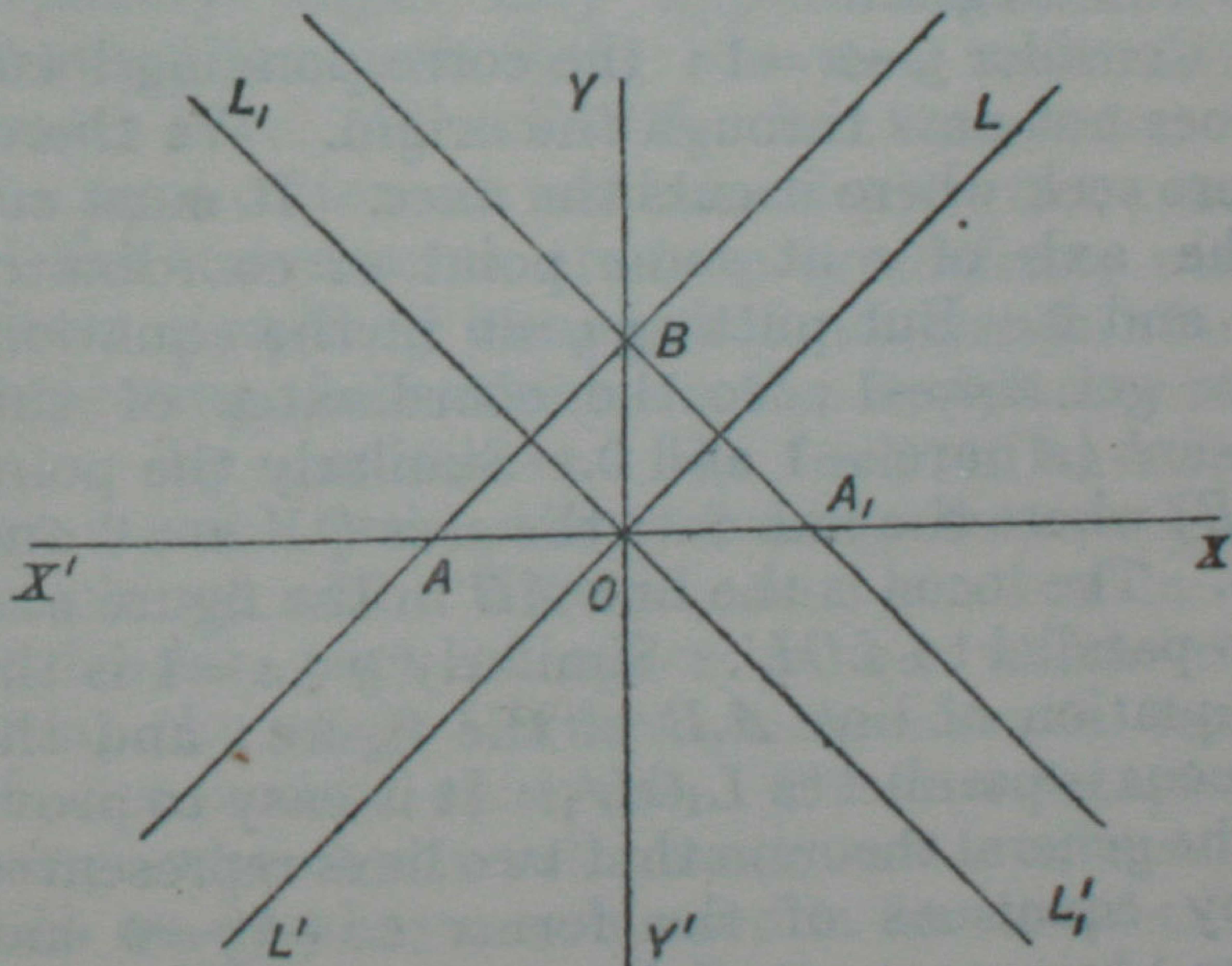


Fig. 14.

Consider $y-x=0$; here the a, b, and c, of the general form have been replaced by -1, 1, and 0 respectively. This line passes through the "origin," O, in the diagram and bisects the angle XOY. It is the line $L'OL$ of the diagram. The fact that it passes through the origin, O, is easily seen by observing that the equation is satisfied by putting $x=0$ and $y=0$ simultaneously, but 0 and 0 are the coordinates of O. In fact it is easy to generalize and to see by the same method that the equation of any line through the origin is of the form $ax+by=0$. The locus of equation $y+x=0$ also passes through the origin and bisects the angle $X'OY$: it is the line $L_1OL'_1$ of the diagram.

Consider $y-x=1$: the corresponding locus does not pass through the origin. We therefore seek where it cuts the axes. It must cut the axis of x at some point of coordinates x and 0. But putting $y=0$ in the equation, we get $x=-1$; so the coordinates of this point (A) are -1 and 0. Similarly the point (B) where the line cuts the axis OY are 0 and 1. The locus is the line AB in the figure and is parallel to LOL'. Similarly $y+x=1$ is the equation of line A_1B of the figure; and the locus is parallel to $L_1OL'_1$. It is easy to prove the general theorem that two lines represented by equations of the forms $ax+by=0$ and $ax+by=c$ are parallel.

The group of loci which we next come upon are sufficiently important to deserve a chapter to themselves. But before going on to them we will dwell a little longer on the main ideas of the subject.

The position of any point P is determined by arbitrarily choosing an origin, O, two axes, OX and OY, at right-angles, and then by noting its coordinates x and y, *i.e.* OM and PM (*cf.* fig. 13). Also, as we have seen in the last chapter, P can be determined by the "vector" OP, where the idea of the vector includes a determinate direction as well as a determinate length. From an abstract mathematical point of view the idea of an arbitrary origin may appear artificial and clumsy, and similarly for the arbitrarily drawn axes, OX and OY. But in relation to the application of mathematics to the event of the Universe we are here symbolizing with direct simplicity the most fundamental fact respecting the outlook on the world afforded to us by our senses. We each of us refer our sensible perceptions of things to an origin which we call "here": our location in a particular part of space round which we group the whole Universe is the essential fact of our bodily existence. We can imagine beings who observe all phenomena in all space with an equal eye, unbiassed in favour of any part. With us it is otherwise, a cat at our feet claims more attention than an earthquake at Cape Horn, or than the destruction of a world in the Milky Way. It is true that in making a common stock of our knowledge with our fellowmen, we have to waive something of the strict egoism of our own individual "here." We substitute "nearly here" for "here"; thus we measure miles from the town hall of the nearest town, or from the capital of the country. In measuring the earth, men of science will put the origin at the earth's centre; astronomers even rise to the extreme altruism of putting their origin inside the sun. But, far as this last origin may be, and even if we go further to some convenient point amid the nearer fixed stars, yet, compared to the immeasurable infinities of space, it remains true that our first procedure in exploring the Universe is to fix upon an origin "nearly here."

Again the relation of the coordinates OM and MP (*i.e.* x and y) to the vector OP is an instance of the famous parallelogram law, as can easily be seen (*cf.* fig. 8) by completing the parallelogram $OMPN$. The idea of the "vector" OP, that is, of a directed magnitude, is the root-idea of physical science. Any moving body has a certain magnitude of velocity in a certain direction, that is to say, its velocity is a directed magnitude, a vector. Again a force has a certain magnitude and has a definite direction. Thus, when in analytical geometry the ideas of the "origin," of "coordinates," and of "vectors" are introduced, we are studying the abstract conceptions which correspond to the fundamental facts of the physical world.

CHAPTER X

CONIC SECTIONS

WHEN the Greek geometers had exhausted, as they thought, the more obvious and interesting properties of figures made up of straight lines and circles, they turned to the study of other curves; and, with their almost infallible instinct for hitting upon things worth thinking about, they chiefly devoted themselves to conic sections, that is, to the curves in which planes would cut the surfaces of circular cones. The man who must have the credit of inventing the study is Menaechmus (born 375 B.C. and died 325 B.C.); he was a pupil of Plato and one of the tutors of Alexander the Great. Alexander, by the by, is a conspicuous example of the advantages of good tuition, for another of his tutors was the philosopher Aristotle. We may suspect that Alexander found Menaechmus rather a dull teacher, for it is related that he asked for the proofs to be made shorter. It was to this request that Menaechmus replied: "In the country there are private and even royal roads, but in geometry there is only one road for all." This reply no doubt was true enough in the sense in which it would have been immediately understood by Alexander. But if Menaechmus thought that his proofs could not be shortened, he was grievously mistaken; and most modern mathematicians would be horribly bored, if they were compelled to study the Greek proofs of the properties of conic sections. Nothing illustrates better the gain in power which is obtained by the introduction of relevant ideas into a science than to observe the progressive shortening of proofs which accompanies the growth of richness in idea. There is a certain type of mathematician who is always rather impatient at delaying over the ideas of a subject: he is anxious at once to get on to the proofs of "important" problems. The history of the science is entirely against him. There are royal roads in science; but those who first tread them are men of genius and not kings.

The way in which conic sections first presented themselves to mathematicians was as follows: think of a cone (*cf.* fig. 15), whose vertex (or point) is V, standing on a circular base STU. For example, a conical shade to an electric light is often an example of such a surface. Now let the "generating" lines which pass through V and lie on the surface be all produced backwards; the result is a double cone, and PQR is another circular cross section on the opposite side of V to the cross section STU. The axis of the cone CVC' passes through all the centres of these circles and is perpendicular to their planes, which are parallel to each other. In the diagram the parts of the curves which are supposed to lie behind the plane of the paper are dotted lines, and the parts on the plane or in front of it are continuous lines. Now suppose this double cone is cut by a plane not perpendicular to the axis CVC', or at least not necessarily perpendicular to it. Then three cases can arise :—

(1) The plane may cut the cone in a closed oval curve, such as $ABA'B'$ which lies entirely on one of the two half-cones. In this case the plane will not meet the other half-cone at all. Such a curve is called an ellipse ; it is an oval curve. A particular case of such a section of the cone is when the plane is perpendicular to the axis CVC', then the section, such as STU or PQR, is a circle. Hence a circle is a particular case of the ellipse.

(2) The plane may be parallel to a tangent plane touching the cone along one of its "generating" lines as for example the plane of the

curve $D_1A_1D_1'$ in the diagram is parallel to the tangent plane touching the cone along the generating line VS; the curve is still confined to one of the half-cones, but it is now not a closed oval curve, it goes on endlessly as long as the generating lines of the half-cone are produced away from the vertex. Such a conic section is called a parabola.

(3) The plane may cut both the half-cones, so that the complete curve consists of two detached portions, or "branches" as they are called, this case is illustrated by the two branches $G_2A_2G_2'$ and $L_2A_2'L_2'$ which together make up the curve. Neither branch is closed, each of them spreading out endlessly as the two half-cones are prolonged away from the vertex. Such a conic section is called a hyperbola.

There are accordingly three types of conic sections, namely, ellipses, parabolas, and hyperbolas. It is easy to see that, in a sense, parabolas are limiting cases lying between ellipses and hyperbolas. They form a more special sort and have to satisfy a more particular condition. These three names are apparently due to Apollonius of Perga (born about 260 B.C., and died about 200 B.C.), who wrote a systematic treatise on conic sections which remained the standard work till the sixteenth century.

It must at once be apparent how awkward

and difficult the investigation of the properties of these curves must have been to the Greek geometers. The curves are plane curves, and yet their investigation involves

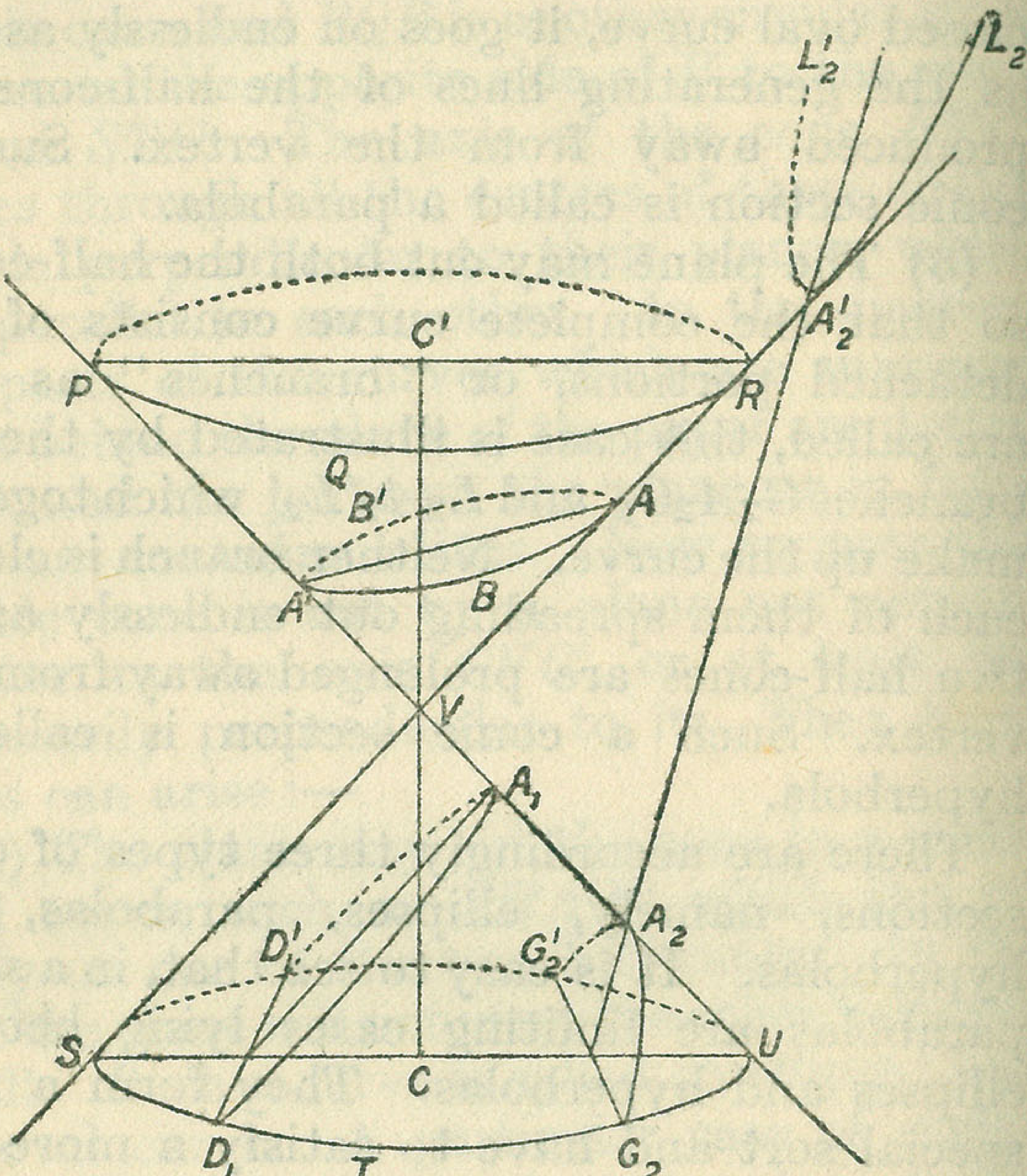


Fig. 15

the drawing in perspective of a solid figure. Thus in the diagram given above we have practically drawn no subsidiary lines and yet the figure is sufficiently complicated. The curves are plane curves, and it seems obvious that we should be able to define them without

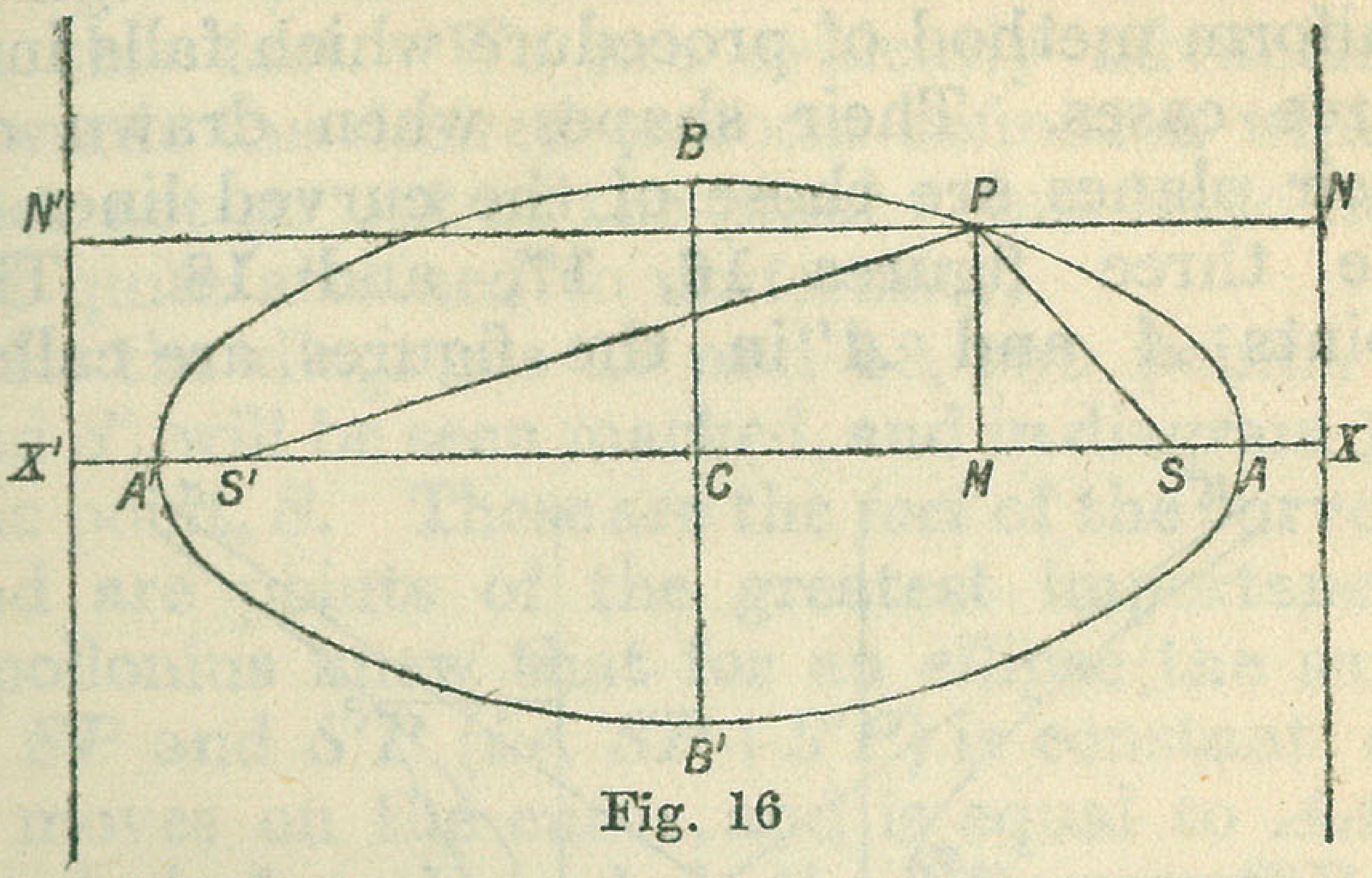


Fig. 16

going beyond the plane into a solid figure. At the same time, just as in the "solid"

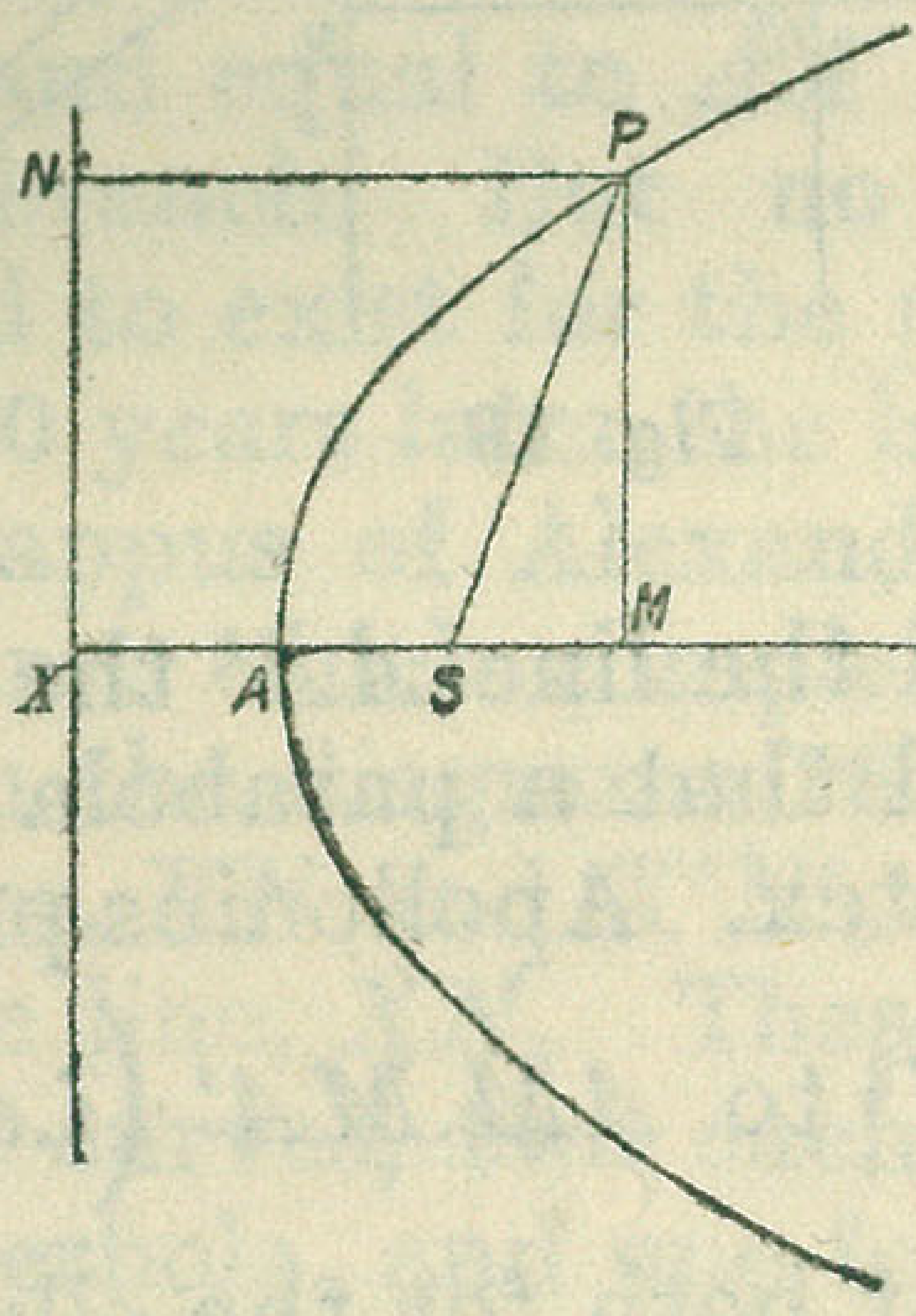


Fig. 17

definition there is one uniform method of definition—namely, the section of a cone by a plane—which yields three cases, so in any "plane" definition there also should be one uniform method of procedure which falls into three cases. Their shapes when drawn on their planes are those of the curved lines in the three figures 16, 17, and 18. The points A and A' in the figures are called

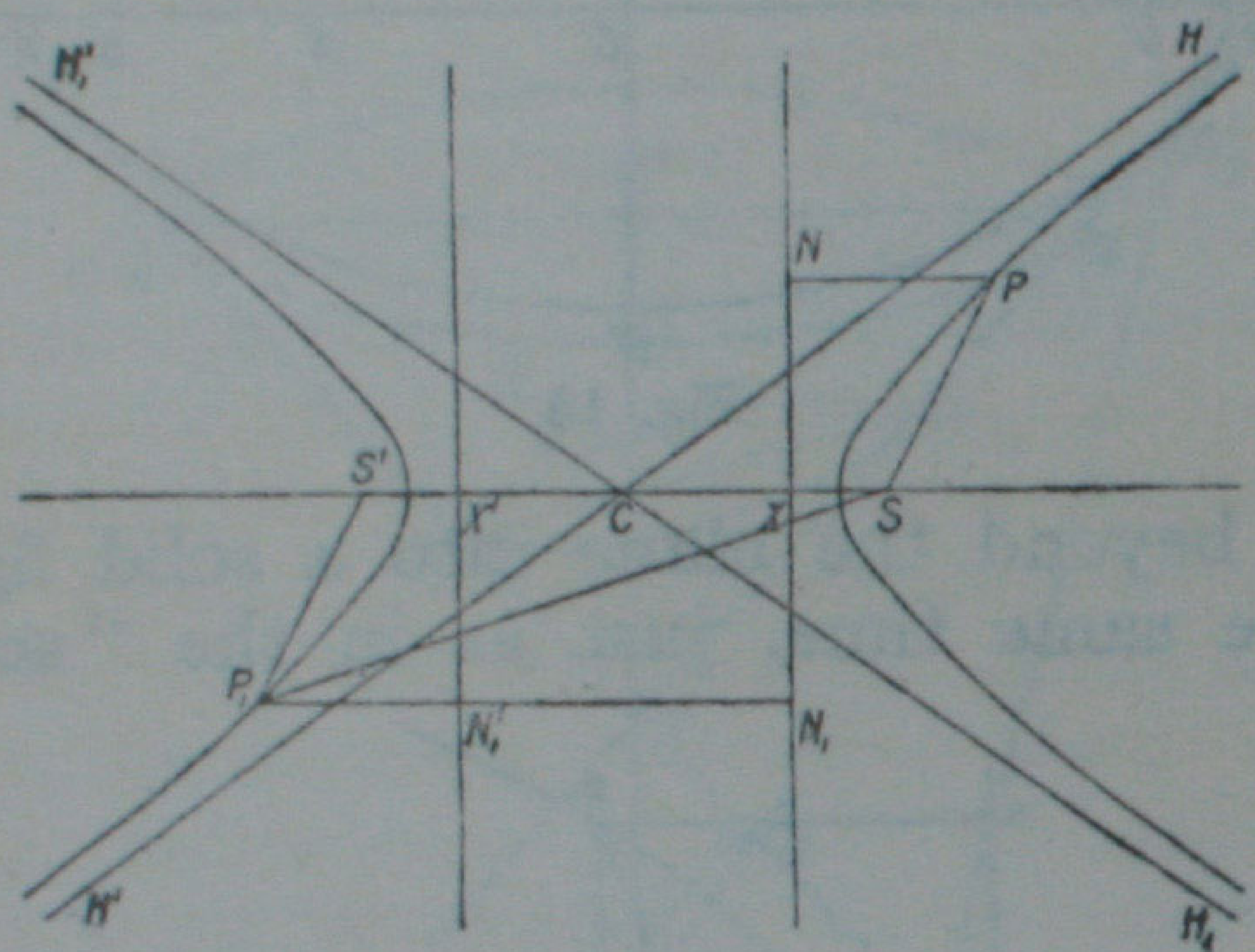


Fig. 18

the vertices and the line AA' the major axis. It will be noted that a parabola (*cf.* fig. 17) has only one vertex. Apollonius proved* that the ratio of PM^2 to $AM.MA'$ $\left(i.e. \dfrac{PM^2}{AM.MA'}\right)$ remains constant both for the ellipse and the hyperbola (figs. 16 and 18), and that the ratio

* Cf. Ball, *loc. cit.*, for this account of Apollonius and Pappus.

of PM^2 to AM is constant for the parabola of fig. 17; and he bases most of his work on this fact. We are evidently advancing towards the desired uniform definition which does not go out of the plane; but have not yet quite attained to uniformity.

In the diagrams 16 and 18, two points, S and S', will be seen marked, and in diagram 17 one point, S. These are the *foci* of the curves, and are points of the greatest importance. Apollonius knew that for an ellipse the sum of SP and $S'P$ (*i.e.* $SP+S'P$) is constant as P moves on the curve, and is equal to AA'. Similarly for a hyperbola the difference $S'P-SP$ is constant, and equal to AA' when P is on one branch, and the difference $SP'-S'P'$ is constant and equal to AA' when P' is on the other branch. But no corresponding point seemed to exist for the parabola.

Finally 500 years later the last great Greek geometer, Pappus of Alexandria, discovered the final secret which completed this line of thought. In the diagrams 16 and 18 will be seen two lines, XN and $X'N'$, and in diagram 17 the single line, XN. These are the directrices of the curves, two each for the ellipse and the hyperbola, and one for the parabola. Each directrix corresponds to its nearer focus. The characteristic property of a focus, S, and its corresponding directrix, XN, for any one of the three types of curve, is that the ratio

SP to PN $\left(i.e. \frac{SP}{PN}\right)$ is constant, where PN is the perpendicular on the directrix from P, and P is any point on the curve. Here we have finally found the desired property of the curves which does not require us to leave the plane, and is stated uniformly for all three curves. For ellipses the ratio* $\frac{SP}{PN}$ is less than 1, for parabolas it is equal to 1, and for hyperbolas it is greater than 1.

When Pappus had finished his investigations, he must have felt that, apart from minor extensions, the subject was practically exhausted; and if he could have foreseen the history of science for more than a thousand years, it would have confirmed his belief. Yet in truth the really fruitful ideas in connection with this branch of mathematics had not yet been even touched on, and no one had guessed their supremely important applications in nature. No more impressive warning can be given to those who would confine knowledge and research to what is apparently useful, than the reflection that conic sections were studied for eighteen hundred years merely as an abstract science, without a thought of any utility other than to satisfy the craving for knowledge on the part of mathematicians, and that then at the end of this long period of abstract study, they

* Cf. Note B, p. 250.

were found to be the necessary key with which to attain the knowledge of one of the most important laws of nature.

Meanwhile the entirely distinct study of astronomy had been going forward. The great Greek astronomer Ptolemy (died 168 A.D.) published his standard treatise on the subject in the University of Alexandria, explaining the apparent motions among the fixed stars of the sun and planets by the conception of the earth at rest and the sun and the planets circling round it. During the next thirteen hundred years the number and the accuracy of the astronomical observations increased, with the result that the description of the motions of the planets on Ptolemy's hypothesis had to be made more and more complicated. Copernicus (born 1473 A.D. and died 1543 A.D.) pointed out that the motions of these heavenly bodies could be explained in a simpler manner if the sun were supposed to rest, and the earth and planets were conceived as moving round it. However, he still thought of these motions as essentially circular, though modified by a set of small corrections arbitrarily superimposed on the primary circular motions. So the matter stood when Kepler was born at Stuttgart in Germany in 1571 A.D. There were two sciences, that of the geometry of conic sections and that of astronomy, both of which had been studied from a remote antiquity without a suspicion of any connection between the two. Kepler was an astronomer, but he was also an able geometer, and on the subject of conic sections had arrived at ideas in advance of his time He is only one of many examples of the falsity of the idea that success in scientific research demands an exclusive absorption in one narrow line of study. Novel ideas are more apt to spring from an unusual assortment of knowledge—not necessarily from vast knowledge, but from a thorough conception of the methods and ideas of distinct lines of thought. It will be remembered that Charles Darwin was helped to arrive at his conception of the law of evolution by reading Malthus' famous *Essay on Population*, a work dealing with a different subject—at least, as it was then thought.

Kepler enunciated three laws of planetary motion, the first two in 1609, and the third ten years later. They are as follows:

(1) The orbits of the planets are ellipses, the sun being in the focus.

(2) As a planet moves in its orbit, the radius vector from the sun to the planet sweeps out equal areas in equal times.

(3) The squares of the periodic times of the several planets are proportional to the cubes of their major axes.

These laws proved to be only a stage towards a more fundamental development of ideas. Newton (born 1642 A.D. and died 1727 A.D.) conceived the idea of universal gravitation, namely, that any two pieces of matter attract each other with a force proportional to the product of their masses and inversely proportional to the square of their distance from each other. This sweeping general law, coupled with the three laws of motion which he put into their final general shape, proved adequate to explain all astronomical phenomena, including Kepler's laws, and has formed the basis of modern physics. Among other things he proved that comets might move in very elongated ellipses, or in parabolas, or in hyperbolas, which are nearly parabolas. The comets which return—such as Halley's comet—must, of course, move in ellipses. But the essential step in the proof of the law of gravitation, and even in the suggestion of its initial conception, was the verification of Kepler's laws connecting the motions of the planets with the theory of conic sections.

From the seventeenth century onwards the abstract theory of the curves has shared in the double renaissance of geometry due to the introduction of coordinate geometry and of projective geometry. In projective geometry the fundamental ideas cluster round the consideration of sets (or pencils, as they are called) of lines passing through a common point (the vertex of the "pencil"). Now (*cf.* fig. 19) if *A*, *B*, *C*, *D*, be any four fixed points on a conic section and P be a variable point on the curve, the pencil of lines *PA*,

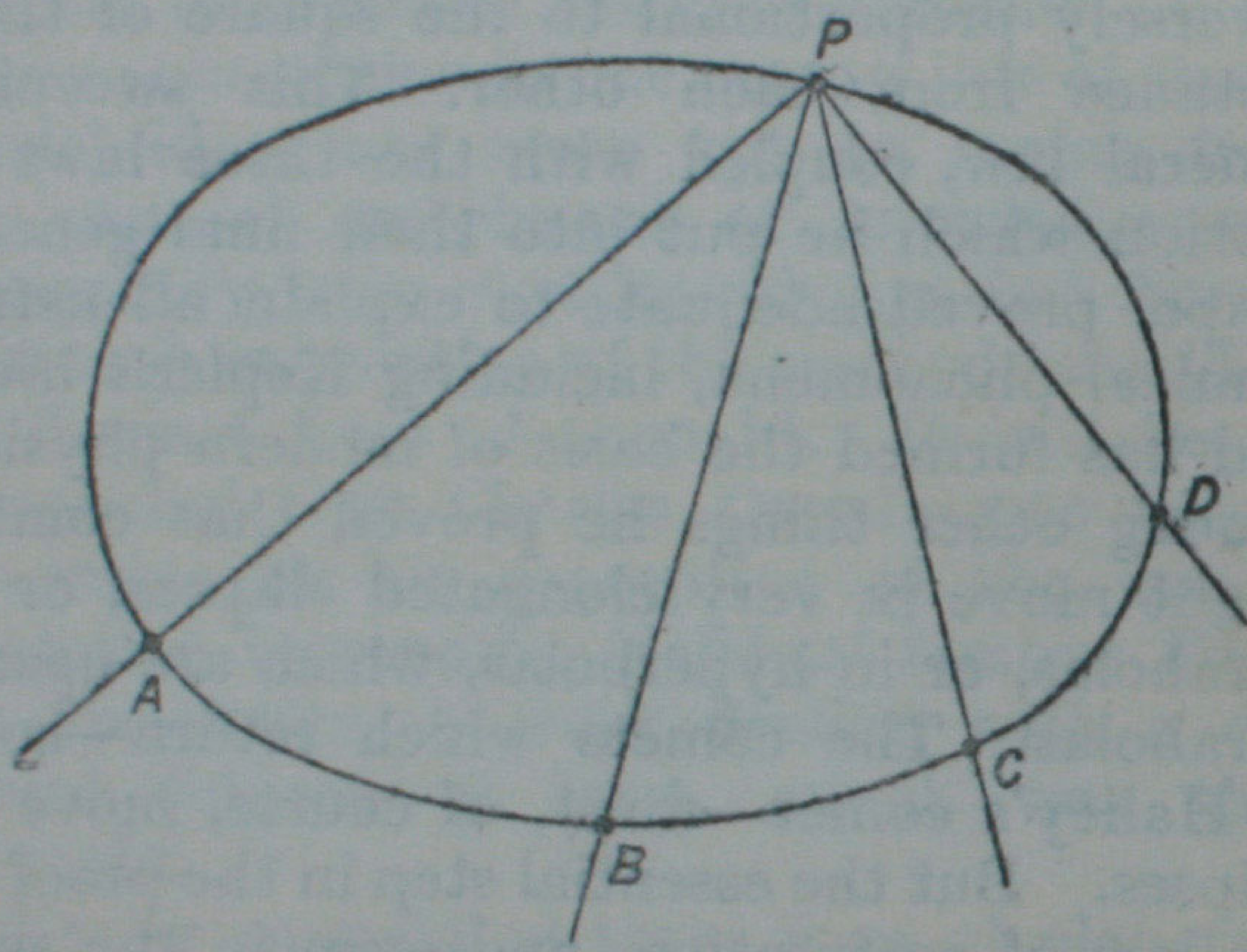


Fig. 19.

PB, *PC*, and *PD*, has a special property, known as the constancy of its cross ratio. It will suffice here to say that cross ratio is a fundamental idea in projective geometry. For projective geometry this is really the definition of the curves, or some analogous property which is really equivalent to it. It will be seen how far in the course of ages of study we have drifted away from the old original idea of the sections of a circular cone. We know now that the Greeks had got hold of a minor property of comparatively slight importance; though by some divine good fortune the curves themselves deserved all the attention which was paid to them. This unimportance of the "section" idea is now marked in ordinary mathematical phraseology by dropping the word from their names. As often as not, they are now named merely "conics" instead of "conic sections."

Finally, we come back to the point at which we left coordinate geometry in the last chapter. We had asked what was the type of *loci* corresponding to the general algebraic form $ax+by=c$, and had found that it was the class of straight lines in the plane. We had seen that every straight line possesses an equation of this form, and that every equation of this form corresponds to a straight line. We now wish to go on to the next general type of algebraic forms. This is evidently to be obtained by introducing terms involving x^2 and xy and y^2. Thus the new general form must be written—

$$ax^2+2hxy+by^2+2gx+2fy+c=0$$

What does this represent? The answer is

that (when it represents any locus) it always represents a conic section, and, furthermore, that the equation of every conic section can always be put into this shape. The discrimination of the particular sorts of conics as given by this form of equation is very easy. It entirely depends upon the consideration of $ab-h^2$, where a, b, and h, are the "constants" as written above. If $ab-h^2$ is a positive number, the curve is an ellipse; if $ab-h^2=0$, the curve is a parabola: and if $ab-h^2$ is a negative number, the curve is a hyperbola.

For example, put $a=b=1$, $h=g=f=0$, $c=-4$. We then get the equation $x^2+y^2-4=0$. It is easy to prove that this is the equation of a circle, whose centre is at the origin, and radius is 2 units of length. Now $ab-h^2$ becomes $1\times1-0^2$, that is, 1, and is therefore positive. Hence the circle is a particular case of an ellipse, as it ought to be. Generalising, the equation of any circle can be put into the form $a(x^2+y^2)+2gx+2fy+c=0$. Hence $ab-h^2$ becomes a^2-0, that is, a^2, which is necessarily positive. Accordingly all circles satisfy the condition for ellipses. The general form of the equation of a parabola is

$$(dx+ey)^2+2gx+2fy+c=0,$$

so that the terms of the second degree, as they are called, can be written as a perfect square. For squaring out, we get

$$d^2x^2+2dexy+e^2y^2+2gx+2fy+c;$$

so that by comparison $a=d^2$, $h=de$, $b=e^2$, and therefore $ab-h^2=d^2e^2-(de)^2=0$. Hence the necessary condition is automatically satisfied. The equation $2xy-4=0$, where $a=b=g=f=0$, $h=1$, $c=-4$, represents a hyperbola. For the condition $ab-h^2$ becomes $0-1^2$, that is, -1, which is negative.

The limitation, introduced by saying that, *when the general equation represents any locus*, it represents a conic section, is necessary, because some particular cases of the general equation represent no real locus. For example $x^2+y^2+1=0$ can be satisfied by no real values of x and y. It is usual to say that the locus is now one composed of imaginary points. But this idea of imaginary points in geometry is really one of great complexity, which we will not now enter into.

Some exceptional cases are included in the general form of the equation which may not be immediately recognized as conic sections. By properly choosing the constants the equation can be made to represent two straight lines. Now two intersecting straight lines may fairly be said to come under the Greek idea of a conic section. For, by referring to the picture of the double cone above, it will be seen that some planes through the vertex, V, will cut the cone in a pair of straight lines intersecting at V. The case of two parallel straight lines can be included by considering a circular cylinder as a particular case of a cone. Then a plane, which cuts it and is parallel to its axis, will cut it in two parallel straight lines. Anyhow, whether or no the ancient Greek would have allowed these special cases to be called conic sections, they are certainly included among the curves represented by the general algebraic form of the second degree. This fact is worth noting; for it is characteristic of modern mathematics to include among general forms all sorts of particular cases which would formerly have received special treatment. This is due to its pursuit of generality.

CHAPTER XI

FUNCTIONS

THE mathematical use of the term function has been adopted also in common life. For example, "His temper is a function of his digestion," uses the term exactly in this mathematical sense. It means that a rule can be assigned which will tell you what his temper will be when you know how his digestion is working. Thus the idea of a "function" is simple enough, we only have to see how it is applied in mathematics to variable numbers. Let us think first of some concrete examples: If a train has been travelling at the rate of twenty miles per hour, the distance (s miles) gone after any number of hours, say t, is given by $s=20\times t$; and s is called a function of t. Also $20\times t$ is the function of t with which s is identical. If John is one year older than Thomas, then, when Thomas is at any age of x years, John's age (y years) is given by $y=x+1$; and y is a function of x, namely, is the function $x+1$.

In these examples t and x are called the "arguments" of the functions in which they appear. Thus t is the argument of the function $20 \times t$, and x is the argument of the function $x+1$. If $s=20 \times t$, and $y=x+1$, then s and y are called the "values" of the functions $20 \times t$ and $x+1$ respectively.

Coming now to the general case, we can define a function in mathematics as a correlation between two variable numbers, called respectively the argument and the value of the function, such that whatever value be assigned to the "argument of the function" the "value of the function" is definitely (*i.e.* uniquely) determined. The converse is not necessarily true, namely, that when the value of the function is determined the argument is also uniquely determined. Other functions of the argument x are $y=x^2$, $y=2x^2+3x+1$, $y=x$, $y=log\ x$, $y=sin\ x$. The last two functions of this group will be readily recognizable by those who understand a little algebra and trigonometry. It is not worth while to delay now for their explanation, as they are merely quoted for the sake of example.

Up to this point, though we have defined what we mean by a function in general, we have only mentioned a series of special functions. But mathematics, true to its general methods of procedure, symbolizes the general idea of any function. It does this by writing

$F(x)$, $f(x)$, $g(x)$, $\phi(x)$, etc., for any function of x, where the argument x is placed in a bracket, and some letter like F, f, g, ϕ, etc., is prefixed to the bracket to stand for the function. This notation has its defects. Thus it obviously clashes with the convention that the single letters are to represent variable numbers; since here F, f, g, ϕ, etc., prefixed to a bracket stand for variable functions. It would be easy to give examples in which we can only trust to common sense and the context to see what is meant. One way of evading the confusion is by using Greek letters (*e.g.* ϕ as above) for functions; another way is to keep to f and F (the initial letter of function) for the functional letter, and, if other variable functions have to be symbolized, to take an adjacent letter like g.

With these explanations and cautions, we write $y=f(x)$, to denote that y is the value of some undetermined function of the argument x; where $f(x)$ may stand for anything such as $x+1$, x^2-2x+1, $\sin x$, $\log x$, or merely for x itself. The essential point is that when x is given, then y is thereby definitely determined. It is important to be quite clear as to the generality of this idea. Thus in $y=f(x)$, we may determine, if we choose, $f(x)$ to mean that when x is an integer, $f(x)$ is zero, and when x has any other value, $f(x)$ is 1. Accordingly, putting $y=f(x)$, with this choice for the meaning of f, y is either 0 or 1 according as the value of x is integral or otherwise. Thus $f(1)=0$, $f(2)=0$, $f(\frac{2}{3})=1$, $f(\sqrt{2})=1$, and so on. This choice for the meaning of $f(x)$ gives a perfectly good function of the argument x according to the general definition of a function.

A function, which after all is only a sort of correlation between two variables, is represented like other correlations by a graph, that is in effect by the methods of coordinate geometry. For example, fig. 2 in Chapter II. is the graph of the function $\frac{1}{v}$ where v is the argument and p the value of the function. In this case the graph is only drawn for positive values of v, which are the only values possessing any meaning for the physical application considered in that chapter. Again in fig. 14 of Chapter IX. the whole length of the line AB, unlimited in both directions, is the graph of the function $x+1$, where x is the argument and y is the value of the function; and in the same figure the unlimited line A_1B is the graph of the function $1-x$, and the line LOL' is the graph of the function x, x being the argument and y the value of the function.

These functions, which are expressed by simple algebraic formulæ, are adapted for representation by graphs. But for some functions this representation would be very misleading without a detailed explanation, or might even be impossible. Thus, consider the function mentioned above, which has the value 1 for all values of its argument x, except those which are integral, *e.g.* except for $x=0$, $x=1$, $x=2$, etc., when it has the value 0. Its appearance on a graph would be that of the straight line ABA' drawn parallel to the

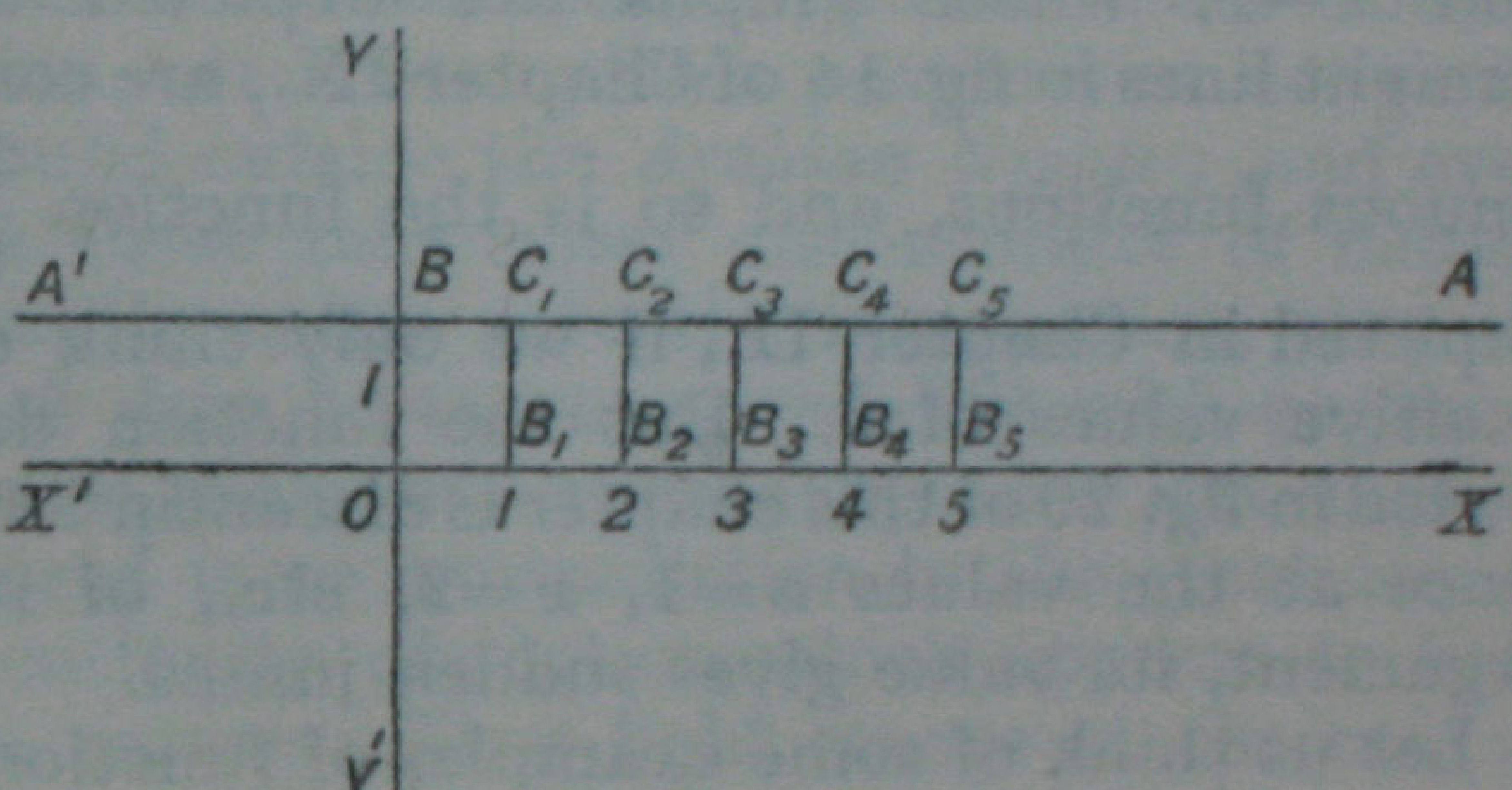


Fig. 20.

axis XOX' at a distance from it of 1 unit of length. But the points, B, C_1, C_2, C_3, C_4, etc., corresponding to the values 0, 1, 2, 3, 4, etc., of the argument x, are to be omitted, and instead of them the points O, B_1, B_2, B_3, B_4, etc., on the axis OX, are to be taken. It is easy to find functions for which the graphical representation is not only inconvenient but impossible. Functions which do not lend themselves to graphs are important in the higher mathematics, but we need not concern ourselves further about them here.

The most important division between functions is that between continuous and discontinuous functions. A function is continuous when its value only alters gradually for gradual alterations of the argument, and is discontinuous when it can alter its value by sudden jumps. Thus the two functions $x+1$ and $1-x$, whose graphs are depicted as straight lines in fig. 14 of Chapter IX., are continuous functions, and so is the function $\frac{1}{v}$, depicted in Chapter II., if we only think of positive values of v. But the function depicted in fig. 20 of this chapter is discontinuous since at the values $x=1$, $x=2$, etc., of its argument, its value gives sudden jumps.

Let us think of some examples of functions presented to us in nature, so as to get into our heads the real bearing of continuity and discontinuity. Consider a train in its journey along a railway line, say from Euston Station, the terminus in London of the London and North-Western Railway. Along the line in order lie the stations of Bletchley and Rugby. Let t be the number of hours which the train has been on its journey from Euston, and s be the number of miles passed over. Then s is a function of t, *i.e.* is the variable value corresponding to the variable argument t.

If we know the circumstances of the train's run, we know s as soon as any special value of t is given. Now, miracles apart, we may confidently assume that s is a continuous function of t. It is impossible to allow for the contingency that we can trace the train continuously from Euston to Bletchley, and that then, without any intervening time, however short, it should appear at Rugby. The idea is too fantastic to enter into our calculation: it contemplates possibilities not to be found outside the *Arabian Nights*; and even in those tales sheer discontinuity of motion hardly enters into the imagination, they do not dare to tax our credulity with anything more than very unusual speed. But unusual speed is no contradiction to the great law of continuity of motion which appears to hold in nature. Thus light moves at the rate of about 190,000 miles per second and comes to us from the sun in seven or eight minutes; but, in spite of this speed, its distance travelled is always a continuous function of the time.

It is not quite so obvious to us that the velocity of a body is invariably a continuous function of the time. Consider the train at any time t: it is moving with some definite velocity, say v miles per hour, where v is zero when the train is at rest in a station and is negative when the train is backing. Now we readily allow that v cannot change its value suddenly for a big, heavy train. The train certainly cannot be running at forty miles per hour from 11.45 a.m. up to noon, and then suddenly, without any lapse of time, commence running at 50 miles per hour. We at once admit that the change of velocity will be a gradual process. But how about sudden blows of adequate magnitude? Suppose two trains collide; or, to take smaller objects, suppose a man kicks a football. It certainly appears to our sense as though the football began suddenly to move. Thus, in the case of velocity our senses do not revolt at the idea of its being a discontinuous function of the time, as they did at the idea of the train being instantaneously transported from Bletchley to Rugby. As a matter of fact, if the laws of motion, with their conception of mass, are true, there is no such thing as discontinuous velocity in nature. Anything that appears to our senses as discontinuous change of velocity must, according to them, be considered to be a case of gradual change which is too quick to be perceptible to us. It would be rash, however, to rush into the generalization that no discontinuous functions are presented to us in nature. A man who, trusting that the mean height of the land above sea-level between London and Paris was a continuous function of the distance from London, walked at night on Shakespeare's Cliff by Dover in contemplation of the Milky Way, would be dead before he had had time to rearrange his ideas as to the necessity of caution in scientific conclusions.

It is very easy to find a discontinuous function, even if we confine ourselves to the

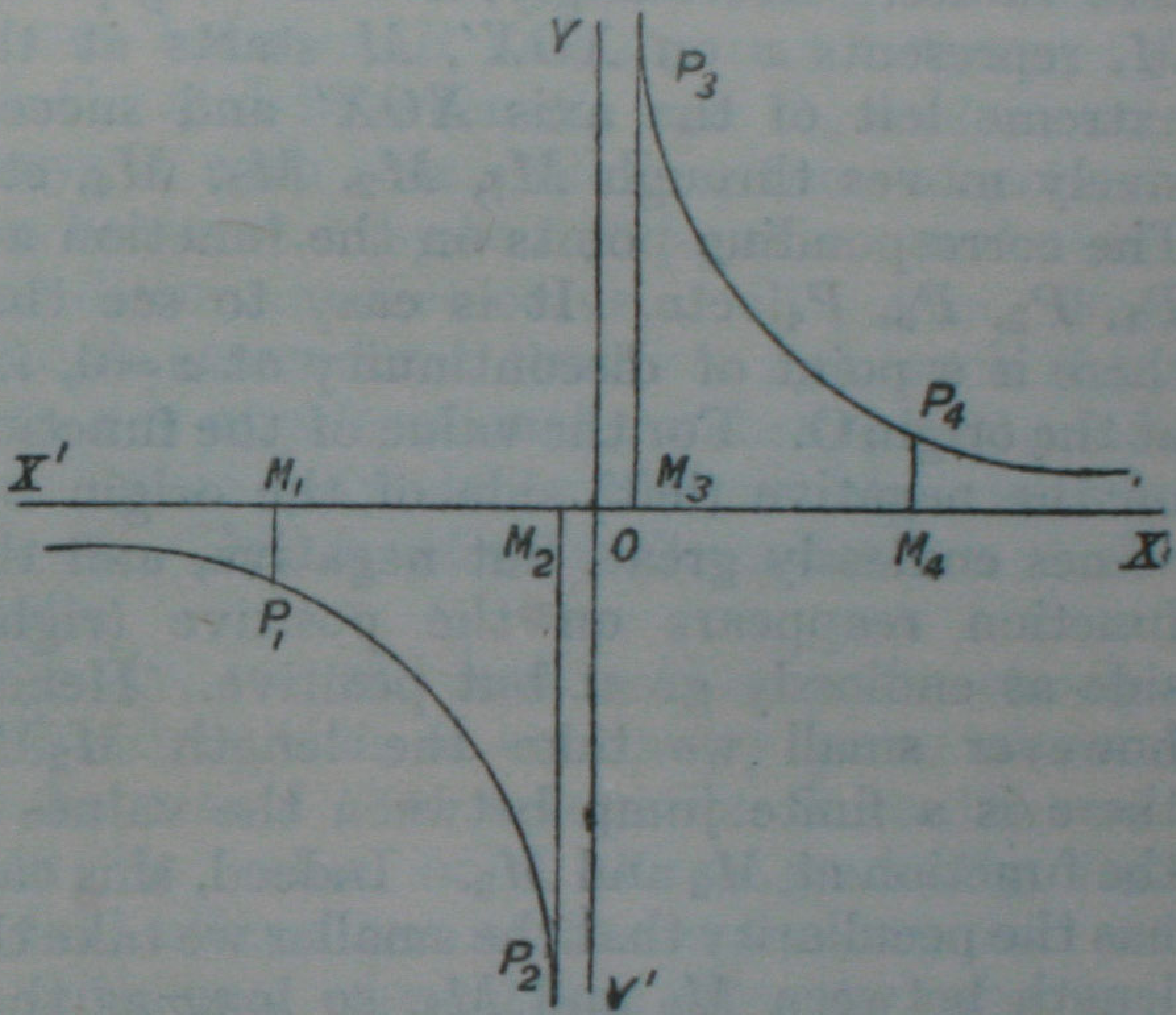


Fig. 21.

simplest of the algebraic formulæ. For example, take the function $y=\frac{1}{x}$, which we have already considered in the form $p=\frac{1}{v}$, where v was confined to positive values. But now let x have any value, positive or negative. The graph of the function is exhibited in fig. 21. Suppose x to change continuously from a large negative value through a numerically decreasing set of negative values up to 0, and thence through the series of increasing positive values. Accordingly, if a moving point, M, represents x on XOX', M starts at the extreme left of the axis XOX' and successively moves through M_1, M_2, M_3, M_4, etc. The corresponding points on the function are P_1, P_2, P_3, P_4, etc. It is easy to see that there is a point of discontinuity at $x=0$, *i.e.* at the origin O. For the value of the function on the negative (left) side of the origin becomes endlessly great, but negative, and the function reappears on the positive (right) side as endlessly great but positive. Hence, however small we take the length M_2M_3, there is a finite jump between the values of the function at M_2 and M_3. Indeed, this case has the peculiarity that the smaller we take the length between M_2 and M_3, so long as they enclose the origin, the bigger is the jump in value of the function between them. This graph brings out, what is also apparent in fig. 20 of this chapter, that for many functions the discontinuities only occur at isolated points, so that by restricting the values of the argument we obtain a continuous function for these remaining values. Thus it is evident from fig. 21 that in $y=\frac{1}{x}$, if we keep to positive values only and exclude the origin, we obtain a continuous function. Similarly the same function, if we keep to negative values only, excluding the origin, is continuous. Again the function which is graphed in fig. 20 is continuous between B and C_1, and between C_1 and C_2, and between C_2 and C_3, and so on, always in each case excluding the end points. It is, however, easy to find functions such that their discontinuities occur at all points. For example, consider a function $f(x)$, such that when x is any fractional number $f(x)=1$, and when x is any incommensurable number $f(x)=2$. This function is discontinuous at all points.

Finally, we will look a little more closely at the definition of continuity given above. We have said that a function is continuous when its value only alters gradually for gradual alterations of the argument, and is discontinuous when it can alter its value by sudden jumps. This is exactly the sort of definition which satisfied our mathematical forefathers and no longer satisfies modern mathematicians. It is worth while to spend some time over it; for when we understand the modern objections to it, we shall have gone a long way towards the understanding of the spirit of modern mathematics. The whole difference between the older and the newer mathematics lies in the fact that vague half-metaphorical terms like "gradually" are no longer tolerated in its exact statements. Modern mathematics will only admit statements and definitions and arguments which exclusively employ the few simple ideas about number and magnitude and variables on which the science is founded. Of two numbers one can be greater or less than the other; and one can be such and such a multiple of the other; but there is no relation of "graduality" between two numbers, and hence the term is inadmissible. Now this may seem at first sight to be great pedantry. To this charge there are two answers. In the first place, during the first half of the nineteenth century it was found by some great mathematicians, especially Abel in Sweden, and Weierstrass in Germany, that large parts of mathematics as enunciated in the old happy-go-lucky manner were simply wrong. Macaulay in his essay on Bacon contrasts the certainty of mathematics with the uncertainty of philosophy; and by way of a rhetorical example he says, "There has been no reaction against Taylor's theorem." He could not have chosen a worse example. For, without having made an examination of English text-books on mathematics contemporary with the publication of this essay, the assumption is a fairly safe one that Taylor's theorem was enunciated and proved wrongly in every one of them. Accordingly, the anxious precision of modern mathematics is necessary for accuracy. In the second place it is necessary for research. It makes for clearness of thought, and thence for boldness of thought and for fertility in trying new combinations of ideas. When the initial statements are vague and slipshod, at every subsequent stage of thought common sense has to step in to limit applications and to explain meanings. Now in creative thought common sense is a bad master. Its sole criterion for judgment is that the new ideas shall look like the old ones. In other words it can only act by suppressing originality.

In working our way towards the precise definition of continuity (as applied to functions) let us consider more closely the statement that there is no relation of "graduality" between numbers. It may be asked, Cannot one number be only slightly greater than another number, or in other words, cannot the difference between the two numbers be small? The whole point is that in the abstract, apart from some arbitrarily assumed application, there is no such thing as a great or a small number. A million miles is a small number of miles for an astronomer investigating the fixed stars, but a million pounds is a large yearly income. Again, one-quarter is a large fraction of one's income to give away in charity, but is a small fraction of it to retain for private use. Examples can be accumulated indefinitely to show that great or small in any absolute sense have no abstract application to numbers. We can say of two numbers that one is greater or smaller than another, but not without specification of particular circumstances that any one number is great or small. Our task therefore is to define continuity without any mention of a "small" or "gradual" change in value of the function.

In order to do this we will give names to some ideas, which will also be useful when we come to consider limits and the differential calculus.

An "interval" of values of the argument x of a function $f(x)$ is all the values lying between some two values of the argument. For example, the interval between $x=1$ and $x=2$ consists of all the values which x can take lying between 1 and 2, *i.e.* it consists of all the real numbers between 1 and 2. But the bounding numbers of an interval need not be integers. An interval of values of the argument *contains* a number a, when a is a member of the interval. For example, the interval between 1 and 2 contains $\frac{3}{2}$, $\frac{5}{3}$, $\frac{7}{4}$, and so on.

A set of numbers approximates to a number a within a *standard* k, when the numerical difference between a and every number of the set is less than k. Here k is the "standard of approximation." Thus the set of numbers 3, 4, 6, 8, approximates to the number 5 within the standard 4. In this case the standard 4 is not the smallest which could have been chosen, the set also approximates to 5 within any of the standards 3·1 or 3·01 or 3·001. Again, the numbers, 3·1, 3·141, 3·1415, 3·14159 approximate to 3·13102 within the standard ·032, and also within the smaller standard ·03103.

These two ideas of an interval and of approximation to a number within a standard are easy enough; their only difficulty is that they look rather trivial. But when combined with the next idea, that of the "neighbourhood" of a number, they form the foundation of modern mathematical reasoning. What do we mean by saying that something is true for a function $f(x)$ in the neighbourhood of the value a of the argument x? It is this fundamental notion which we have now got to make precise.

The values of a function $f(x)$ are said to possess a characteristic in the "neighbourhood of a" when some interval can be found, which (i) contains the number a not as an end-point, and (ii) is such that every value of the function for arguments, other than a, lying within that interval possesses the characteristic. The value $f(a)$ of the function for the argument a may or may not possess the characteristic. Nothing is decided on this point by statements about the *neighbourhood* of a.

For example, suppose we take the particular function x^2. Now *in the neighbourhood of* 2, the values of x^2 are less than 5. For we can find an interval, *e.g.* from 1 to 2·1, which (i) contains 2 not as an end-point, and (ii) is such that, for values of x lying within it, x^2 is less than 5.

Now, combining the preceding ideas we know what is meant by saying that *in the neighbourhood of a* the function $f(x)$ approximates to c within the *standard k*. It means that some interval can be found which (i) includes a not as an end-point, and (ii) is such that all values of $f(x)$, where x lies in the interval and is not a, differ from c by less than k. For example, in the neighbourhood of 2, the function $\sqrt{x}$ approximates to 1·41425 within the standard ·0001. This is true because the square root of 1·99996164 is 1·4142 and the square root of 2·00024449 is 1·4143; hence for values of x lying in the interval 1·99996164 to 2·00024449, which contains 2 not as an end-point, the values of the function $\sqrt{x}$ all lie between 1·4142 and 1·4143, and they therefore all differ from 1·41425 by less than ·0001. In this case we can, if we like, fix a smaller standard of approximation, namely ·000051 or ·0000501. Again, to take another example, in the neighbourhood of 2 the function x^2 approximates to 4 within the standard ·5. For $(1{\cdot}9)^2=3{\cdot}61$ and $(2{\cdot}1)^2=4{\cdot}41$, and thus the required interval 1·9 to 2·1, containing 2 not as an end-point, has been found. This example brings out the fact that statements about a function $f(x)$ in the neighbourhood of a number a are distinct from statements about the value of $f(x)$ when $x=a$. The production of an *interval*, throughout which the statement is true, is required. Thus the mere fact that $2^2=4$ does not by itself justify us in saying that in the *neighbourhood* of 2 the function x^2 is equal to 4. This statement would be untrue, because no interval can be produced with the required property. Also, the fact that $2^2=4$ does not by itself justify us in saying that in the *neighbourhood* of 2 the function x^2 approximates to 4 within the standard ·5; although as a matter of fact, the statement has just been proved to be true.

If we understand the preceding ideas, we understand the foundations of modern mathematics. We shall recur to analogous ideas in the chapter on Series, and again in the chapter on the Differential Calculus.

Meanwhile, we are now prepared to define "continuous functions." A function $f(x)$ is "continuous" at a value a of its argument, when in the neighbourhood of a its values approximate to $f(a)$ (*i.e.* to its value at a) within *every* standard of approximation.

This means that, whatever standard k be chosen, in the neighbourhood of a $f(x)$ approximates to $f(a)$ within the standard k. For example, x^2 is continuous at the value 2 of its argument, x, because however k be chosen we can always find an interval, which (i) contains 2 not as an end-point, and (ii) is such that the values of x^2 for arguments lying within it approximate to 4 (*i.e.* 2^2) within the standard k. Thus, suppose we choose the standard $\cdot 1$; now $(1\cdot 999)^2 = 3\cdot 996001$, and $(2\cdot 01)^2 = 4\cdot 0401$, and both these numbers differ from 4 by less than $\cdot 1$. Hence, within the interval $1\cdot 999$ to $2\cdot 01$ the values of x^2 approximate to 4 within the standard $\cdot 1$. Similarly an interval can be produced for any other standard which we like to try.

Take the example of the railway train. Its velocity is continuous as it passes the signal box, if whatever velocity you like to assign (say one-millionth of a mile per hour) an interval of time can be found extending before and after the instant of passing, such that at all instants within it the train's velocity differs from that with which the train passed the box by less than one-millionth of a mile per hour; and the same is true whatever other velocity be mentioned in the place of one-millionth of a mile per hour.

CHAPTER XII

PERIODICITY IN NATURE

THE whole life of Nature is dominated by the existence of periodic events, that is, by the existence of successive events so analogous to each other that, without any straining of language, they may be termed recurrences of the same event. The rotation of the earth produces the successive days. It is true that each day is different from the preceding days, however abstractly we define the meaning of a day, so as to exclude casual phenomena. But with a sufficiently abstract definition of a day, the distinction in properties between two days becomes faint and remote from practical interest; and each day may then be conceived as a recurrence of the phenomenon of one rotation of the earth. Again the path of the earth round the sun leads to the yearly recurrence of the seasons, and imposes another periodicity on all the operations of nature. Another less fundamental periodicity is provided by the phases of the moon. In modern civilized life, with its artificial light, these phases are of slight importance, but in ancient times, in climates where the days are burning and the skies clear, human life was apparently largely influenced by the existence of moonlight. Accordingly our divisions into weeks and months, with their religious associations, have spread over the European races from Syria and Mesopotamia, though independent observances following the moon's phases are found amongst most nations. It is, however, through the tides, and not through its phases of light and darkness, that the moon's periodicity has chiefly influenced the history of the earth.

Our bodily life is essentially periodic. It is dominated by the beatings of the heart, and the recurrence of breathing. The presupposition of periodicity is indeed fundamental to our very conception of life. We cannot imagine a course of nature in which, as events progressed, we should be unable to say: "This has happened before." The whole conception of experience as a guide to conduct would be absent. Men would always find themselves in new situations possessing no substratum of identity with anything in past history. The very means of measuring time as a quantity would be absent. Events might still be recognized as occurring in a series, so that some were earlier and others later. But we now go beyond this bare recognition. We can not only say that three events, *A*, *B*, *C*, occurred in this order, so that *A* came before *B*, and *B* before *C*; but also we can say that the length of time between the occurrences of *A* and *B* was twice as long as that between *B* and *C*. Now, quantity of time is essentially dependent on observing the number of natural recurrences which have intervened. We may say that the length of time between *A* and *B* was so many days, or so many months, or so many years, according to the type of recurrence to which we wish to appeal. Indeed, at the beginning of civilization, these three modes of measuring time were really distinct. It has been one of the first tasks of science among civilized or semi-civilized nations, to fuse them into one coherent measure. The full extent of this task must be grasped. It is necessary to determine, not merely what number of days (*e.g.* 365·25 . . .) go to some one year, but also previously to determine that the same number of days do go to the successive years. We can imagine a world in which periodicities exist, but such that no two are coherent. In some years there might be 200 days and in others 350. The determination of the broad general consistency of the more important periodicities was the first step in natural science. This consistency arises from no abstract intuitive law of thought; it is merely an observed fact of nature guaranteed by experience. Indeed, so far is it from being a necessary law, that it is not even exactly true There are divergencies in every case. For some instances these divergencies are easily observed and are therefore immediately apparent. In other cases it requires the most refined observations and astronomical accuracy to make them apparent. Broadly speaking, all recurrences depending on living beings, such as the beatings of the heart, are subject in comparison with other recurrences to rapid variations. The great stable obvious recurrences—stable in the sense of mutually agreeing with great accuracy—are those depending on the motion of the earth as a whole, and on similar motions of the heavenly bodies.

We therefore assume that these astronomical recurrences mark out equal intervals of time. But how are we to deal with their discrepancies which the refined observations of astronomy detect? Apparently we are reduced to the arbitrary assumption that one or other of these sets of phenomena marks out equal times—*e.g.* that either all days are of equal length, or that all years are of equal length. This is not so: some assumptions must be made, but the assumption which underlies the whole procedure of the astronomers in determining the measure of time is that the laws of motion are exactly verified.

Before explaining how this is done, it is interesting to observe that this relegation of the determination of the measure of time to the astronomers arises (as has been said) from the stable consistency of the recurrences with which they deal. If such a superior consistency had been noted among the recurrences characteristic of the human body, we should naturally have looked to the doctors of medicine for the regulation of our clocks.

In considering how the laws of motion come into the matter, note that two inconsistent modes of measuring time will yield different variations of velocity to the same body. For example, suppose we define an hour as one twenty-fourth of a day, and take the case of a train running uniformly for two hours at the rate of twenty miles per hour. Now take a grossly inconsistent measure of time, and suppose that it makes the first hour to be twice as long as the second hour. Then, according to this other measure of duration, the time of the train's run is divided into two parts, during each of which it has traversed the same distance, namely, twenty miles; but the duration of the first part is twice as long as that of the second part. Hence the velocity of the train has not been uniform, and on the average the velocity during the second period is twice that during the first period. Thus the question as to