XXVII.

ON THE CANONICAL FORM AND DISSECTION OF A RIEMANN'S SURFACE*.

THE object of this Note is to assist students of the theory of complex functions, by proving the chief propositions about Riemann's surfaces in a concise and elementary manner. To this end I assume only certain results of Puiseux, which are put together at the outset.

I.

Puiseux's Theory of an n-valued Function.

If two variables s and z are connected by an equation of the form $f(s, z), = (s, 1)^n (z, 1)^m, = 0$, each is said to be an algebraic function of the other. Regarding z as a complex quantity $x + iy$, we represent its value by the point whose co-ordinates are x, y, on a certain plane. To every point in this plane belongs one value of z, and consequently, in general, n values of s, which are the roots of the equation $f = 0$. The points of the plane may be divided into those at which the n values of s are distinct, and those at which two or more of them are equal. The latter points are finite in number, and correspond to the roots of the equation which is got by equating to zero the discriminant of f in regard to s. If the roots of this equation are distinct, there are $2(n-1)m$ such points, because the discriminant of the

* [From the *Proceedings of the London Mathematical Society*, Vol. VIII., No. 122, pp. 292—304.]

equation of the nth order in s is of degree $2(n-1)$ in the coefficients, and these coefficients are of the order m in z. But a point at which r values of s become equal corresponds to an $(r-1)$-fold root of the discriminant-equation.

Let us now consider an arbitrary point O of the plane [fig. 31], corresponding to a value z_0 of z, which is not a root of the discriminant-equation. Then the equation $f(s, z_0)=0$ will give n different values for s, which we may call $s_1, s_2 \dots s_n$. If we move along any path from the point O to another point P of the plane, the value of z will change continuously, and each of the quantities $s_1, s_2 \dots s_n$ will also change continuously. If therefore the path OP does not go through a point where two values become equal, these n quantities will be distinct all the way, and each of the n values of s at P will belong to a definite one of the values of s at O. But if the path goes through such a point, two or more of the n quantities will become equal and then diverge again, so that it will be impossible after that to distinguish them so as to say which of these belongs to a particular one of the values at the point O. We cannot always avoid this difficulty by going *round* the point, for it is found that the values at P to which the values at O correspond may depend upon the path OP, so that the correspondence is different for a path which goes to the right of the point and for a path which goes to the left of it. When this is the case, the point is called a branch-point. Suppose that, when we go from O to A, the two values p and q of s at O approach one another and become equal at A; then it is found that the value at P which represents p when we go along the path OBP may represent q when we go along OCP, and *vice versâ*. So that, if we travel along $OBPCO$, round the point A and back to O, the values p and q will change continuously into one another. If more than two values are equal at A, the corresponding values at O may be cyclically interchanged by a path going round A. We shall assume, however, that only two values become equal at each branch-point; and, moreover, that no branch-point is at an infinite distance*.

* Roots of the discriminant-equation which are not branch-points correspond to double points on the curve $f(s, z)=0$. Such points behave, in regard to

A path going along any line from O to very near A, then round A in a very small circle, and then back to O along the same line, will be called a *loop*.

If we start from O and go round any closed curve not including any branch-points, the n values of s at O will be restored in the same order. For the path may be gradually shrunk into a point without crossing any branch-points, so that no two of the n values can become confused at any point of it. The same thing is true if the closed path includes *all* the branch-points. Suppose it a large circle through O; then it may be gradually increased till it coincide with the tangent at O, then curved over on the other side, and shrunk up into a point; and during the whole process the n values will be distinct at every point of the path.

We shall now go on to shew that this n-valued function, which we have spread out upon a single plane, may be represented as a *one*-valued function on a surface consisting of n infinite plane sheets, supposed to lie indefinitely near together, and to cross into one another along certain lines. This surface is called a RIEMANN'S surface; we shall demonstrate its existence at the same time that we shew how to construct it in the most convenient form.

II.

Construction of the Riemann's Surface.—Lüroth's Theorem.

Draw loops from O [fig. 32] to all the branch-points, and let the first, A, interchange the values p and q. If we go round all the loops successively, starting with the value p at O, we must, as we have seen, come back to that value; but this may happen before we have used all the loops. Let B be the first branch-point after going round which the value p is restored. Draw a line from A to B cutting all the loops which alter p, but none of the others. Then, if we go round any of the

the function s, like two coincident branch-points belonging to the same pair of values, and they have no influence on the connection of the different values of s.

branch-points between A and B without crossing the line AB or going round any other branch-points, we shall not alter the value p.

Suppose that A interchanges pq, B interchanges ps, and that the branch-points between A and B are 1, 2, 3, 4, interchanging respectively qr, rs, hk, pl. The value q must in fact be changed into the value p through a longer or shorter series of values; the loops interchanging hk and pl are put in as examples. Now if we go round 4 by the dotted loop passing round outside A, the effect is the same as going in succession round A, 1, 2, 3, 4, 3, 2, 1, A. By the time we have gone round A, 1, 2, 3, we cannot have the value p, for that is first restored by B; and we cannot have the value l, for then 4 would restore the value p. Hence we have some value which is not altered by the loop to 4; and consequently, when we retrace our path, we shall come back to the value p.

Next, let us draw a loop to B which passes within the line AB, but goes round all the included branch-points, as in the figure. The effect of this loop will be to change q into p; for it is the same thing as going round 1, 2, 3, B, 3, 2, 1. Now the effect of 1, 2, 3, B is to change q into p, and this p is not altered in coming back because all the branch-points which alter p are outside the line AB.

Suppose then that all the branch-points of this group which alter p are connected with O by loops going round A, so that they no longer alter p; and that B is connected with O by the loop just described, so that no branch-points are contained in the triangle AOB.

Starting now from this new loop OB, with the value p, let us go round all the loops as before from left to right. We know that when all the loops have been gone round, ending with OA, the value p must be restored. If it is not restored before we have gone round OA, we must draw a line BA cutting all the loops which change the value p but none of the others. But if the value p is restored before we have gone round OA, say after going round OC; then we must draw a new loop to C, going round all the branch-points between A and C except those which change the value p. This new loop will, by our previous reasoning, change p into q. Hence, if the value p is restored before we have gone round OA, we can make a new loop OC which changes p into q; and this comes next to OB. To those branch-points whose loops have been cut by this new loop we must draw new loops going round to the right of C, so as not to cut OC. The figure comes then into this form [fig. 33], containing

(1) Loops to the left of OA which do not change the value of p, like the dotted loop $O4$ in the previous figure;

(2) Three consecutive loops OA, OB, OC which change p into q;

(3) Loops to the right of OC which may or may not change p.

If now we start with the loop OC and proceed to the right, the value p must be restored *before* we have gone round OA; for, starting with OA and going all round, we must restore the value p in the end. Let p then be restored by OD; and draw a line CD cutting all those loops which change p, but none of the others. Replace the loops which change p by new ones going round between B and C; and replace OD by a new loop going outside all the branch-points whose loops do not alter p. The figure now consists of these elements:

(1) Two triangles AOB, COD, containing no branch-points, and such that the loops OA, OB, OC, OD interchange p and q;

(2) Loops between OB and OC which do not change p;

(3) Unknown loops between OD and OA.

About these unknown loops we may make three suppositions.

First, suppose that none of them change p. Then the value p cannot be altered by any closed curve starting from O and returning to it which does not cut either of the lines AB, CD.

Secondly, suppose that some of these loops change p, but that, when we start with the loop OD and go round to the right, the value p is first restored by OA or OC. (It is clear that it cannot be first restored by OB, because the two loops OA, OB, taken together, make no change in any value; nor by any loop between OB and OC, for none of them change p.) Then we must join D with A by a line cutting all the loops which change p, but no others; and B with C by a line cutting none of the loops between OB and OC. In that case the value p cannot be altered by any closed curve starting from O and returning to it which does not cut either of the lines BC, DA.

Thirdly, suppose that the value p is restored *before* we come to OA, say at OE. Then we must proceed as before, finding a new line EF which shall have the properties of AB or CD. The figure will then consist of three triangles AOB, COD, EOF, containing no branch-points, and such that the loops OA, OB, OC, OD, OE, OF interchange p and q; loops between OB and OC, and between OD and OE, which do not change p; and unknown loops between OF and OA.

It is clear that this process must ultimately stop, and then we shall be left with a finite number of lines such that, if we start from O, follow any continuous path, and come back again, without crossing any of these lines, we shall not alter the value p. The lines are either AB, CD, EF, &c., or else they are BC, DE, &c.; in either case the loops $OA, OB, \ldots$ interchange p and q.

It follows that, if we take an infinite plane sheet and cut it through along these lines, we may consider a single value of the function s to be attached to every point of the sheet in such a way that this value varies continuously when we move about continuously in the sheet; but there will be different values on the two sides of any cut—namely, we must attach to every point P of the sheet that value of s which changes continuously into p when we go from P to O without crossing any of the cuts. There is only one such value; for if two different paths from O to P gave different values at P, it would be possible to change the value p by means of a closed curve returning to O; and this we have proved not to be the case.

When the lines cut through are AB, CD, ..., the triangles AOB, COD, ... contain no branch-points; but when the lines are BC, DE, ..., the triangles BOC, DOE do in general contain branch-points. We may, however, draw new loops to C, E, ... so as to exclude these branch-points, and the new loops will still change p into q. For no closed curve going round B and C

so as not to cut BC can change the value p, by what we have already proved; but the loop OB changes p into q, therefore OC must change q into p.

We shall assume then that the cuts are AB, CD, ..., and that the triangles AOB, COD, ... contain no branch-points.

Now let us deal with the value q at O in the same way as we have dealt with the value p. It is first to be observed that a path going round one or more of the lines AB makes no change in *any* value at O; so that, if we agree never to cross these lines, we may leave the branch-points A, B, ... entirely out of consideration.

This being so, let us take a loop which changes q into some other value, say r. There must be such a loop, if the function is more than two-valued; for otherwise the values p, q would form a two-valued algebraic function of z, and the expression $f(s, z)$ would have a factor of the second degree in s.

Starting then with this loop, we may proceed in exactly the same way as before, and draw lines $A'B'$, $C'D'$, ... such that a closed curve, starting from O and coming back to it without cutting any of these lines or any of the previously drawn lines, will not alter the value q. Moreover, we shall have drawn loops OA', OB', ..., each of which changes q into r, and such that the triangles $A'OB'$, $C'OD'$, ... contain no branch-points. And since our previous triangles AOB, COD, ... contained no branch-points, it will not have been necessary to cut through them in drawing the new lines $A'B'$, $C'D'$, ...

We shall now speak of the first set of lines AB, CD,... as the lines (pq), and of the second set as the lines (qr).

Let us take two infinite plane sheets, cut them both through along the lines (pq), but only the second one along the lines (qr). To every point of the *first* sheet we will suppose attached that value of s which is arrived at by continuous change of the value p at O; and to every point of the *second*, that value which is arrived at by continuous change of the value q at O.

In each sheet there will be a finite difference in the values on the two sides of each of the cuts (pq); but the value on one side in the upper sheet will be equal to the value on the other side in the lower sheet. At the cut AB, for example, the value continuous with p on the side next to O is equal to the value continuous with q on the side remote from O; because a path taken round A or B from O and back again changes the value p continuously into the value q.

Thus, if we take p, q to denote values at the cut continuous with p, q at O, they will be situated as in the figure [fig. 34], which represents a section across AB perpendicular to the two sheets. If then we make the two sheets cross one another along the lines p, q, as here represented [fig. 35], then these two values will be continuously distributed on the double-sheeted surface so formed.

We may now continue the process with the value r. We must first find a loop which changes r into some other value, say t, and then proceed as before, taking care not to cross the lines qr. (We may cross the lines pq as often as we please, provided that we have not previously crossed the lines qr; for these lines can have no effect upon r unless it has been previously changed into q.) Thus we shall draw lines rt such that the value r cannot be altered by a closed curve not cutting the lines qr or rt, and having their extremities joined to O by loops which change r into t. If we take, then, a third sheet, cut it through along the lines qr and rt, and then join it crosswise to our second sheet along the lines qr; the three values pqr may be continuously distributed on this three-sheeted surface.

By proceeding in this way it is clear that we shall construct an n-sheeted surface, the sheets of which are connected chainwise by cross lines, so that the first is connected only with the second, the second with the third, and so on; but there is no direct connection except between consecutive sheets. And the n values of the function may now be attached to the points of this surface, so that one value only belongs to each point, and that in moving this point about on the surface the value belonging to it always changes continuously. Thus, if we start from a given point of the surface (on a given sheet), and travel by any path so as to come back to the same point (on the same sheet), we shall in all cases return to the former value of the function s.

The theorem that the Riemann's surface may be so constructed that the sheets are only connected *chainwise*—*i. e.*, so that there are no cross-lines except between consecutive sheets —is due to Dr. Lüroth.

III.

Clebsch's Theorem.

All the links between successive sheets except the last may be made to consist of one cross-line only.

First, we shall prove that, if there are two or more lines (pq), one of them may be converted into a line (qr).

The original position of the two lines (pq) and the line (qr) is drawn in fig. [36]. If we move the line qr, keeping, of course, its ends fixed, the effect is to interchange the sheets QR in the area over which it moves; so that, by passing it over the line (pq) on the right, we change this into a line (pr). The position is then as in fig. [37]. If now we pass the remaining line (pq) over this line (pr), we change it into a line (qr); thus we are left with two lines (qr) and one line (pq). [Fig. 38.]

In this way we may convert all but one of the lines (pq) into lines (qr). Then we may convert all but one of the lines (qr) into lines (rs); and so on. Then the first $n-1$ sheets will be connected chainwise by one cross-line each, and the last two by all the remaining cross-lines.

The Riemann's surface is now said to be in its canonical form.

The process of transformation may be made clearer by looking at a section of the three sheets by a plane perpendicular to them cutting the lines pq, qr, pq [figs. 39, 40, 41].

IV.

Transformation of the Riemann's Surface.

The Riemann's surface now consists of n infinite plane sheets, such that the sheet 1 is connected with 2 by a single cross-line, 2 with 3 by another cross-line, and so on; but $(n-1)$

with (n) by a number of cross-lines which we shall call $p+1$. Thus the whole number of cross-lines is $n-2+p+1=n+p-1$. If w is the number of branch-points, this is twice the number of cross-lines, or $w=2(n+p-1)$. Hence $p=\frac{1}{2}w-n+1$.

Let now this n-fold plane be inverted in regard to any point outside it, so that it becomes an n-fold sphere passing through the point. Any two successive sheets of the sphere will be connected by one cross-line, except the two outside sheets, which are connected by $p+1$ cross-lines.

To every point of this n-sheeted spherical surface will correspond one value of the function s, namely, that which belongs to the corresponding point upon the n-fold plane. As for the centre of inversion, it is to be regarded as n distinct points upon the several sheets, corresponding to the n values of s when $z=\infty$.

We shall now prove that this n-fold spherical surface can be transformed without tearing into the surface of a body with p holes in it.

First, suppose we have only two sheets, connected by a single cross-line which joins the branch-points AB. Let the figure [42] represent a section by the plane which bisects AB at right angles.

Suppose each hemisphere of the inner sheet to be moved across the plane of the great circle containing AB (indicated by the dotted line in the figure), so that the points m, n change places. In this process the two hemispheres will have to penetrate and cross each other; but this may be supposed to take place without altering the continuity of either. Each point may be supposed to move on a straight line perpendicular to the dotted plane, till it coincides with what was its reflexion in regard to that plane. The effect on the cross-line will be to change it from the form drawn in fig. [42] to that drawn in fig. [43]; instead of the two sheets crossing along the line, each of them will be doubled under it. The result is that, if we now look down on the double sphere from a point vertically over the line AB, we shall see a spherical shell with a hole in it, in the form of a slit along the line AB [fig. 44]. Conceive the spherical shell to be made of india-rubber or some more elastic substance ;

then by mere stretching, without tearing, the slit may be opened out until the shell takes the form of a flat plate; that is, of a body with *no* holes in it.

Next, consider a two-sheeted spherical surface with $p + 1$ cross-lines, and suppose them all arranged along the same great circle; which may obviously be done by stretching, without tearing, the surface. Let this great circle be the one represented by the dotted line in figs. [42] and [43]. Then we may apply to the inner sheet the same process as before; viz., we may interchange the two hemispheres into which the sheet is divided by the dotted plane. The effect is to convert all the cross-lines into slits or holes in a spherical shell; and we have supposed that there are $p + 1$ of these. One of the slits may be stretched out in the same way as AB was before, so as to convert the spherical shell into a flat plate; but in this flat plate there will remain p holes. A double sphere with $p + 1$ crossing lines is thus converted, without tearing, into the surface of a body with p holes in it.

Lastly, suppose that the inner sheet of this two-sheeted sphere is connected by one cross-line with a third inside sheet, the third sheet by one cross-line with a fourth inside it, and so on, until there are n sheets. Let the inner sheet of all be reflected in regard to the plane of the great circle through its crossing line, so that it makes with the sheet next to it a spherical shell with one hole in it. Then, without tearing, the inner sheet may be shrunk up until it merely covers over this hole. The same process may now be applied to shrink up the second sheet into the third, and so on, until we are left with only the two outside sheets connected by $p + 1$ cross-lines. These, however, as we have seen, may be converted, without tearing, into the surface of a body with p holes in it. Hence the proposition follows, that *an n-sheeted Riemann's surface with w branch-points may be transformed, without tearing, into the surface of a body with p,* $= \frac{1}{2}w - n + 1$, *holes in it.*

V.

The Number of Irreducible Circuits.

A closed curve drawn on a surface is called a *circuit*. If it is possible to move a circuit continuously on the surface until it shrinks up into a point, the circuit is called *reducible*; otherwise it is *irreducible*. In general there is a finite number of irreducible circuits on a closed surface which are *independent*, that is, no one of which can be made by continuous motion to coincide with a path made out of the others. All other irreducible circuits can then be expressed as compounds of these independent ones. For example, on the surface of a ring (*i.e.*, of a body with one hole through it) there are two independent irreducible circuits; one *round the hole*, as *abc* [fig. 45], and one *through the hole*, as *ade*. If a circuit goes neither round the hole nor through the hole, it can be shrunk up into a point. If it cannot be so shrunk up, it must go a certain number of times round or through the hole or both, that is, it may be made up of circuits like *abc* and *ade*.

In the same way we may see that, on the surface of a body having p holes through it, there are $2p$ independent irreducible circuits; one *round* each hole, and one *through* each hole. For simplicity consider the case $p = 3$. We suppose the body in the form to which we reduced the Riemann's surface, namely, that of a flat plate, represented by figs. [46] and [47], in which A, B, C are the holes. The circuits through each hole are so drawn as to connect the hole directly with the outer rim, like the circuit which is drawn through the hole A. A circuit passing through *two* holes, as B, C [fig. 46], may be moved continuously till it consists of two circuits going through the two holes separately. Similarly, a circuit round two or more holes, as B, C [fig. 47], may be pinched at various points until it is made up of circuits round the separate holes. Such a circuit as *abcd* [fig. 46] may be moved into the form *abcd* [fig. 47], in which it consists of two circuits going through the hole A, but in opposite directions. On this account it may be called a *nugatory* circuit.

VI.

The Canonical Dissection.

Suppose now that it is desired to cut through the Riemann's surface in such a way that it shall still hang together, but that it shall no longer be possible to draw an irreducible circuit upon it. This we may do if we successively prevent the different kinds of irreducible circuits considered in the last section. To prevent the possibility of going *round* any hole, we must cut the surface along a circuit which goes *through* the hole. To prevent the passage *through* a hole, we must cut through a circuit which goes *round* a hole.

Let us make sections a_1, a_2, a_3 [figs. 48, 49] round the holes, and b_1, b_2, b_3 through the holes. Then we shall have prevented the drawing of any irreducible circuits except nugatory ones, like *abcd* in the previous figures. To prevent these also, we may cut the surface along the line c_1 which goes from p to q, that is, from a point on b_2 to a point on b_3, and along the line c_2 which goes from q to r, that is, from a point on b_3 to a point on b_1. We must not cut from r to p also, for then we should divide the surface into two separate parts. We may now open out the upper and under portions of the surface in fig. [48], until it assumes the form of fig. [49]. It then becomes obvious that all our cuts form a continuous line, which is now the boundary of the surface, and is made up of the pieces (beginning at p and going round to the right) $c_1\, b_3\, a_3\, b_3\, c_2\, b_1\, a_1\, b_1\, a_1\, b_1\, c_2\, b_3\, a_3\, b_3\, c_1$. Moreover, it is a matter of intuition that no irreducible contour can now be drawn on the surface.

This system of cuts is called a *canonical dissection* of the surface. In the general case it consists of p cuts a going round the holes, p cuts b going through them, and $p-1$ cuts c joining b_2 to b_3, b_3 to b_4, ... b_p to b_1, but not b_1 to b_2. The cuts c may, if we like, join the a-cuts together, or generally they may join the systems (ab) together, a *system* meaning an a-cut and a b-cut belonging to the same hole. In fact, the c-cuts are only of importance as completing the single boundary of the surface, and so enabling us to see that no irreducible circuit is any longer possible.

It only remains to translate this result so that it may be applicable to the original form of the Riemann's surface, viz., an n-fold plane. We shall do this in the case $p = 2$, which will sufficiently explain the general case. We have now two sheets connected by three cross-lines mn, pq, rs [fig. 50]. One of these must be chosen to represent the outer rim of our flat plate; the other two will then correspond to the holes in it. Let mn, pq represent the holes, and rs the outer rim; lines in the upper sheet shall be drawn in full, and lines in the lower sheet shall be dotted. Then we must first make cuts a_1, a_2, which go round the holes mn, pq; these may lie entirely in the upper sheet. Next we must make cuts b_1, b_2, which connect the holes respectively with the outer rim rs. These cuts lie partly in the upper sheet, where they intersect the cuts a, and partly in the under sheet. Lastly, we must connect the system $a_1\, b_1$ with the system $a_2\, b_2$ by the cut c; this is drawn in the figure from b_1 to b_2 in the under sheet. It is impossible to draw an irreducible circuit on the two-fold plane when it is thus dissected*.

In general, we have proved that in the n-sheeted Riemann's surface which represents the function s determined by the equation $f(s, z) = 0$, there are $p + 1$ cross-lines such that if one be taken to represent the rim, and the rest holes, of a flat plate, the surface may be dissected into one on which no irreducible contour is possible by the following process:—Cut the surface along curves a each of which goes round one of the cross-lines taken to represent holes, on one of the sheets of the surface which cross at that line. Connect each of these lines with the one taken to represent the rim by a cut b along a closed curve which crosses each of the two cross-lines once. Then connect the systems (ab) chainwise by $p - 1$ cuts c.

* It is to be understood that a circuit is *reducible* when all parts of it can be continuously moved away to infinity without crossing any branch-point; because in this theory infinity counts as a single point.

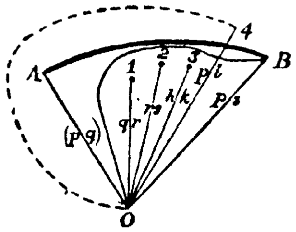

Fig. 32.

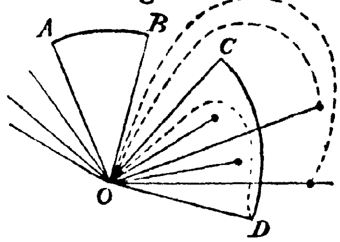

Fig. 33.

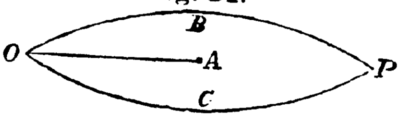

Fig. 31.

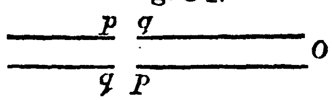

Fig. 34.

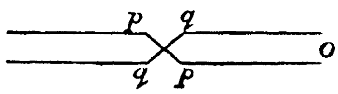

Fig. 35.

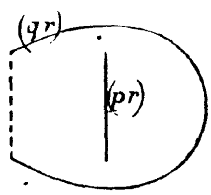

Fig. 37.

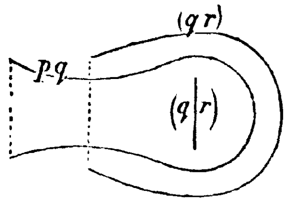

Fig. 38.

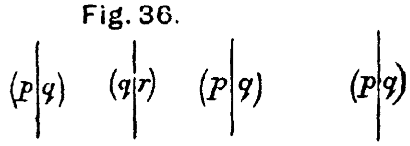

Fig. 36.

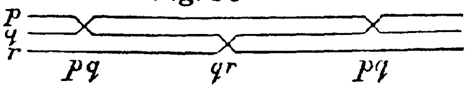

Fig. 39.

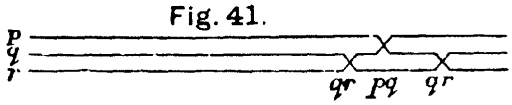

Fig. 41.

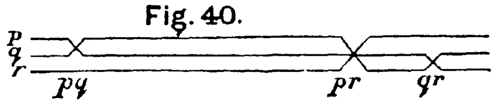

Fig. 40.

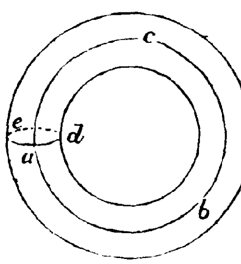

Fig. 45.

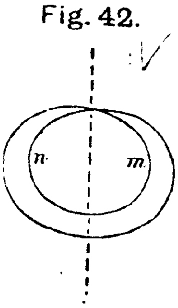

Fig. 42.

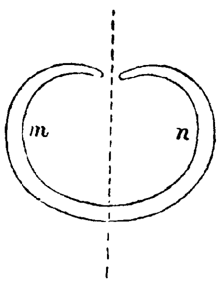

Fig. 43.

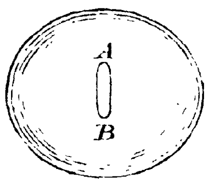

Fig. 44.

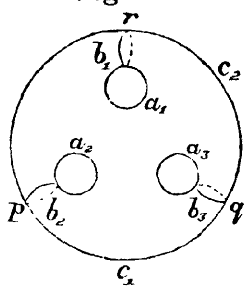

Fig. 48.

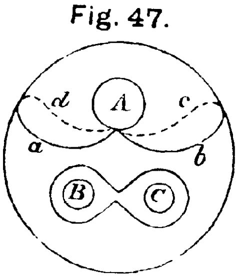

Fig. 47.

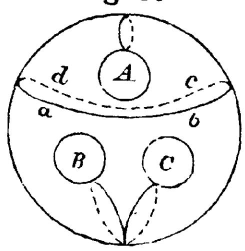

Fig. 46.

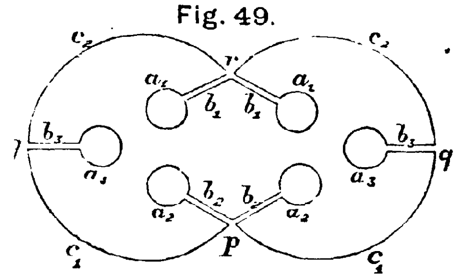

Fig. 49.

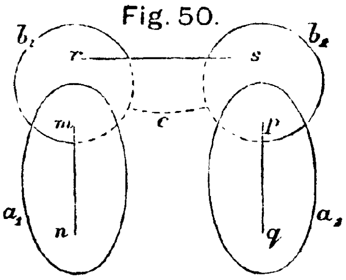

Fig. 50.

face p. 254.