AN INTRODUCTION TO TOPOLOGY

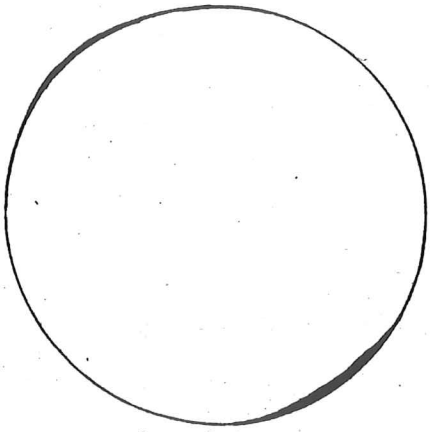
The classification theorem for surfaces.

by

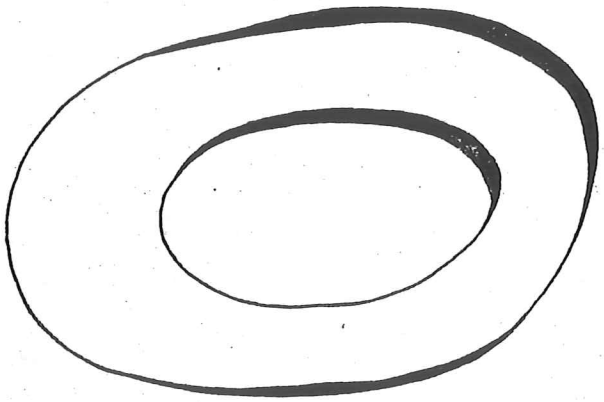
E.C. Zeeman

with diagrams by Fiona.

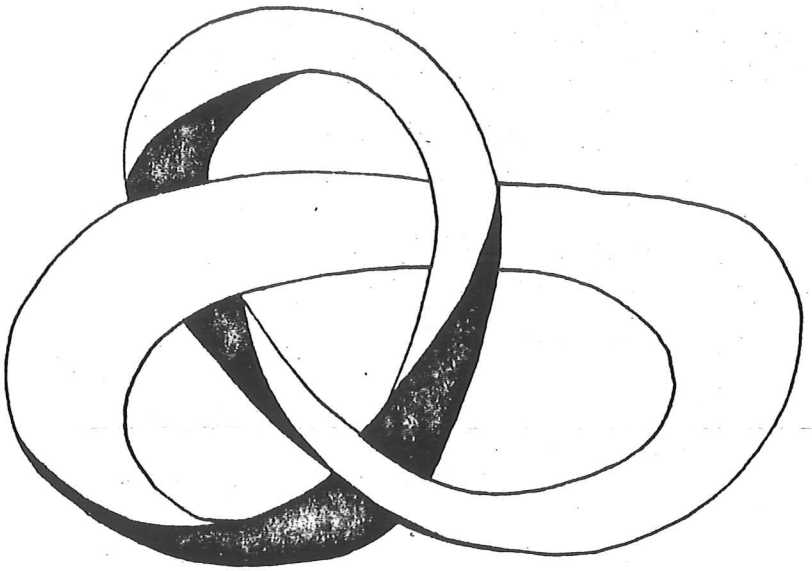
Mathematics Institute,
University of Warwick,
Coventry.



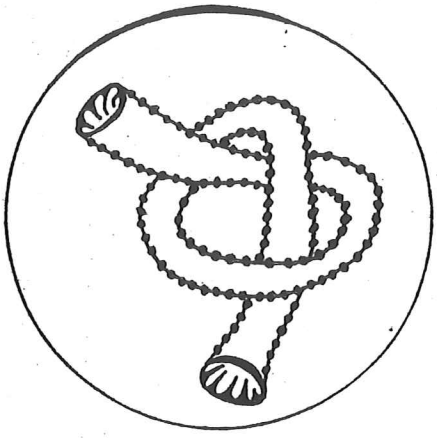
1. Sphere



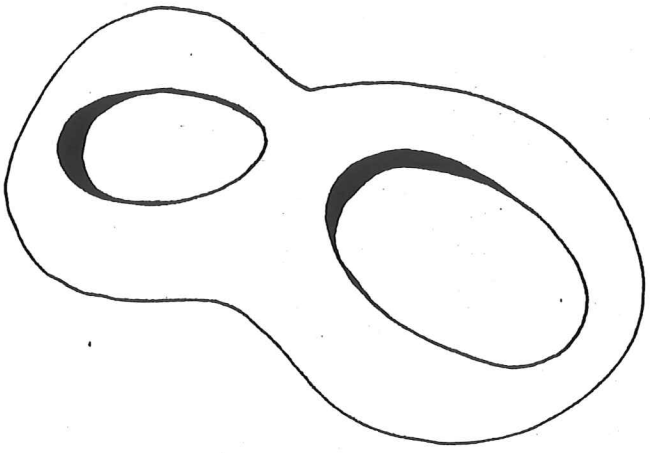
2. Torus (or inner tube).



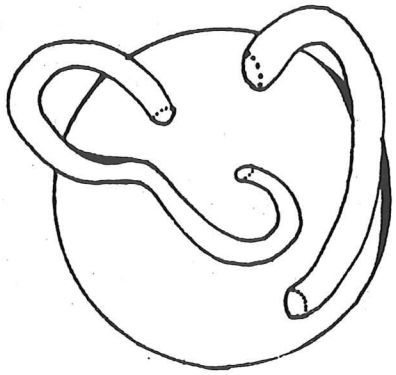
3. Knotted torus



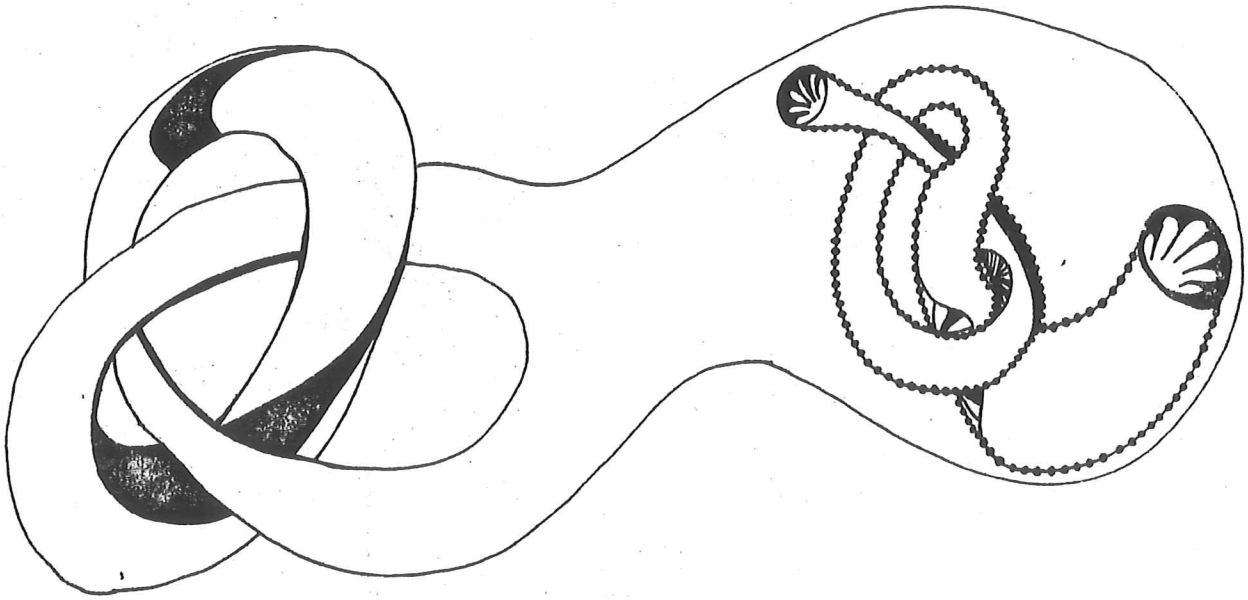
4. Sphere with knotted hole bored through it



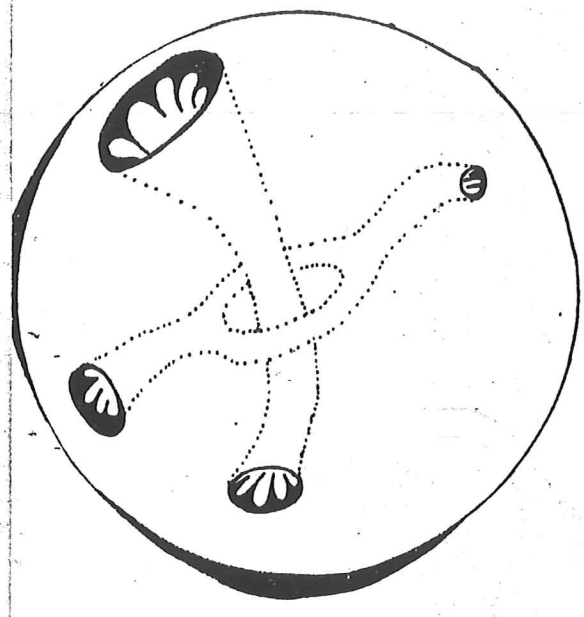
5. Pretzel



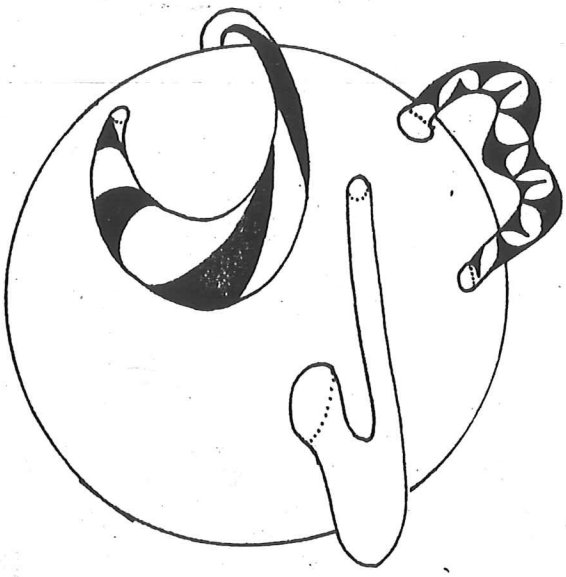
6. Sphere with two handles sewn on.



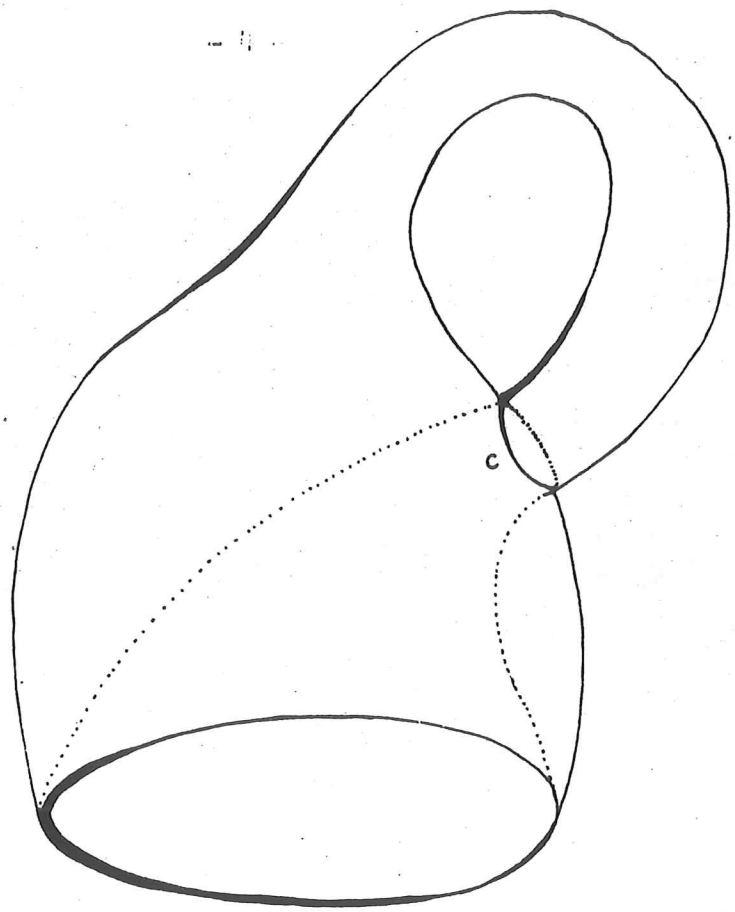
7. Knotted pretzel.



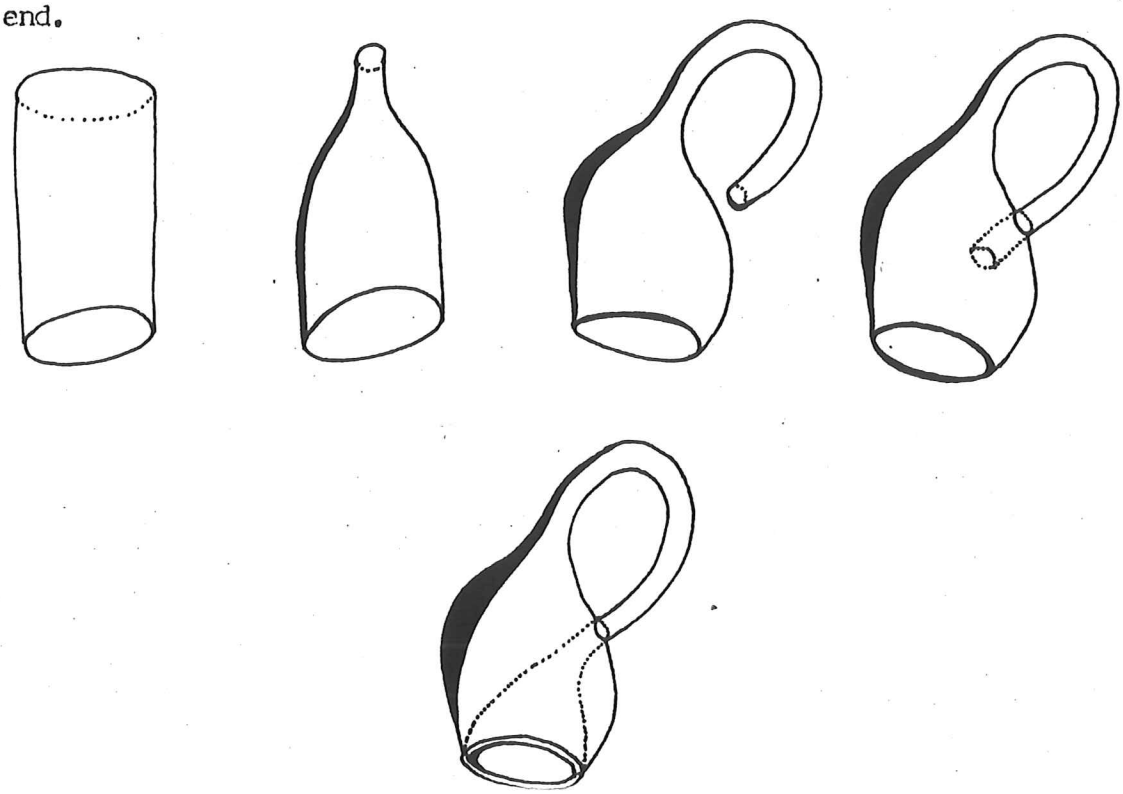
8. Sphere with two holes bored through it, and one of the holes threaded through a hole in the other hole.



9. Sphere with three handles sewn on.



10. Klein bottle. Notice that this surface, unlike the others, intersects itself in the circle C. The Klein bottle can be formed by taking a cylinder, narrowing one end, bending it round, poking it through the side, widening it again, and sewing it onto the other end.



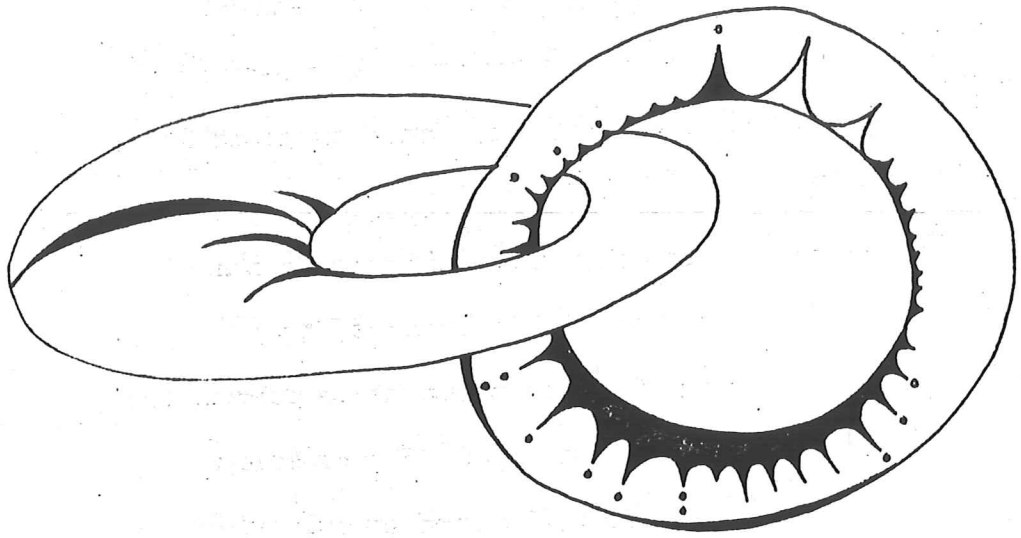
Notice that all the examples above have three properties, which we shall explain

(i) connected

(ii) closed

(iii) triangulable.

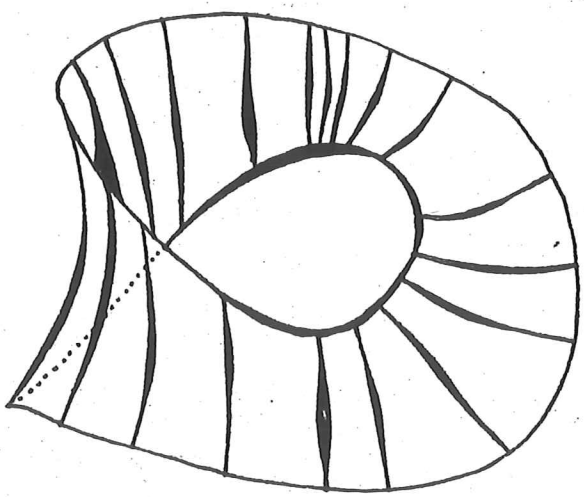
(i) Connected means that the surface is all in one piece. An equivalent definition is that any two points of the surface can be joined by a path in the surface. An example of a surface that is not connected is a pair of tori (possibly linked).



(ii) Closed* means there is no boundary or rim. Examples of surfaces that are not closed are:

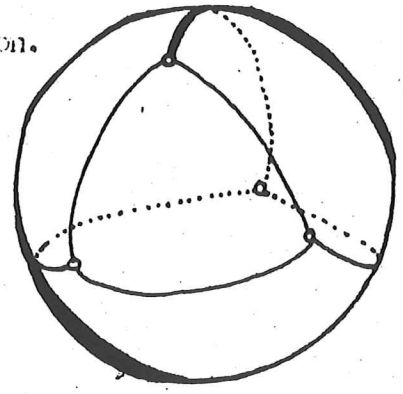
---

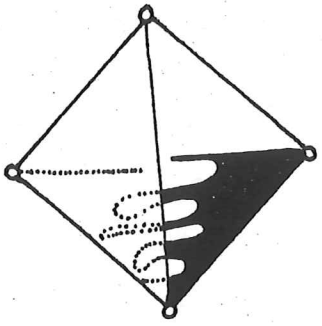
* This usage of the word "closed" is quite different from the usage "open and closed sets" that occurs at the beginning of analytic topology books.



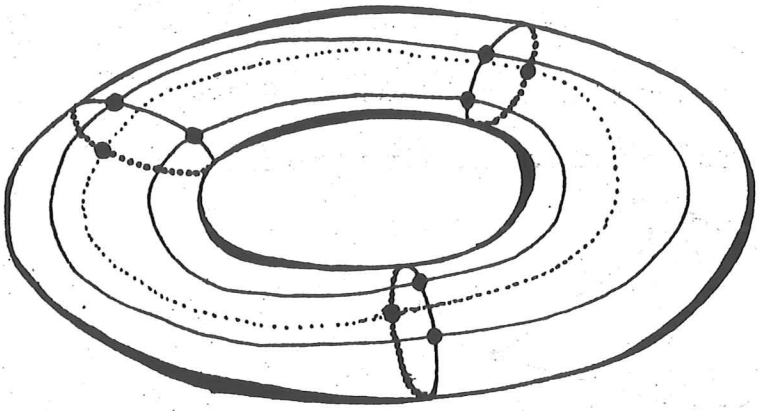
A Möbius strip is formed by taking a strip and sewing the ends together the "wrong way". Notice that a Möbius strip is one-sided in the sense that an ant starting on one side and crawling round it would find itself on the other side.

(iii) Triangulable means that we can chop the surface up into a finite number of vertices, edges and faces. Of course if the surface is curved then the edges and faces also have to be curved, but we can make a model in which they appear straight. For example we can triangulate the sphere with 4 vertices, 6 curved edges, and 4 curved triangles, so that the corresponding straight model is a tetrahedron.

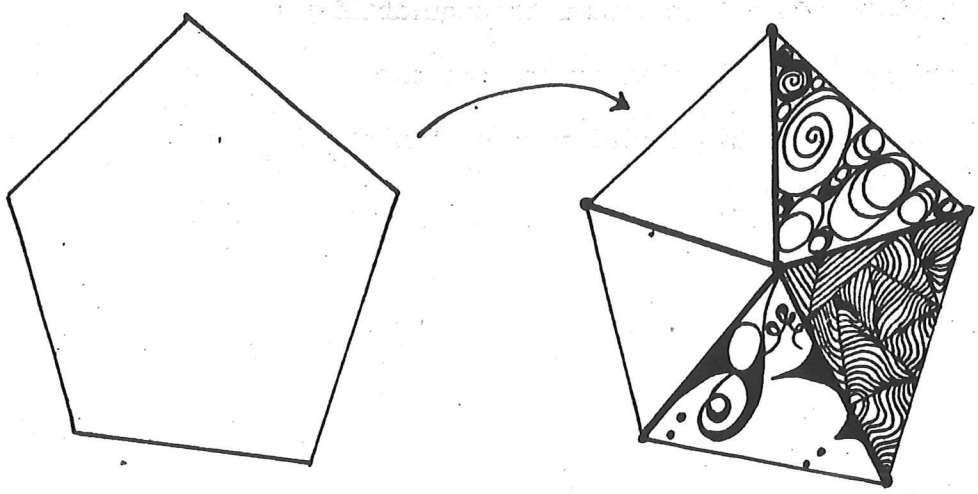




There are lots of ways to chop up a sphere and we could use a million tiny triangles if we wanted to; the main thing is that it can be done in some way. Sometimes it is easier first to chop up a surface into polygons rather than triangles. For example it is easy to chop a torus into 9 vertices, 18 edges and 9 squares.

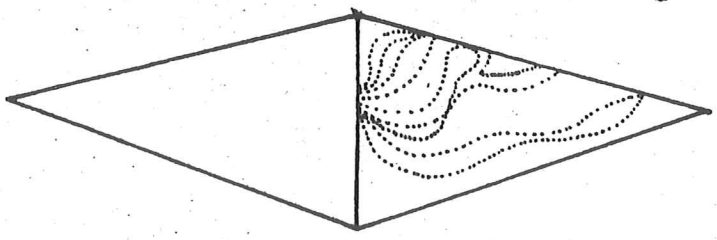


If we can chop a surface into polygons, then we can also chop it into triangles by putting an extra vertex into each polygon:

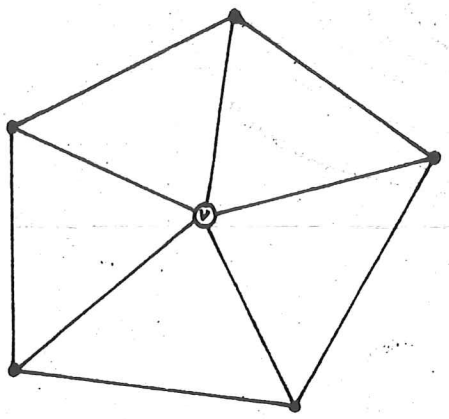


We call a chopping up into triangles a triangulation, and any triangulation of the surfaces has the two properties:

(1) Any edge is the edge of exactly two triangles.



(2) Any vertex, v, is the vertex of at least three triangles, and all the triangles having v as vertex fit round into a cycle.



Our intuition tells us correctly that any surface can be triangulated, but the proof of this fact requires considerable analytic topology, well beyond the scope of this paper (see Reference 4), and so we shall be content to assume triangulability. The great advantage of a triangulation is that it reduces our task of classification to a finite combinatorial problem, which we can then tackle with finite mathematics.

Therefore from now on we shall assume that all our surfaces are

(i) connected

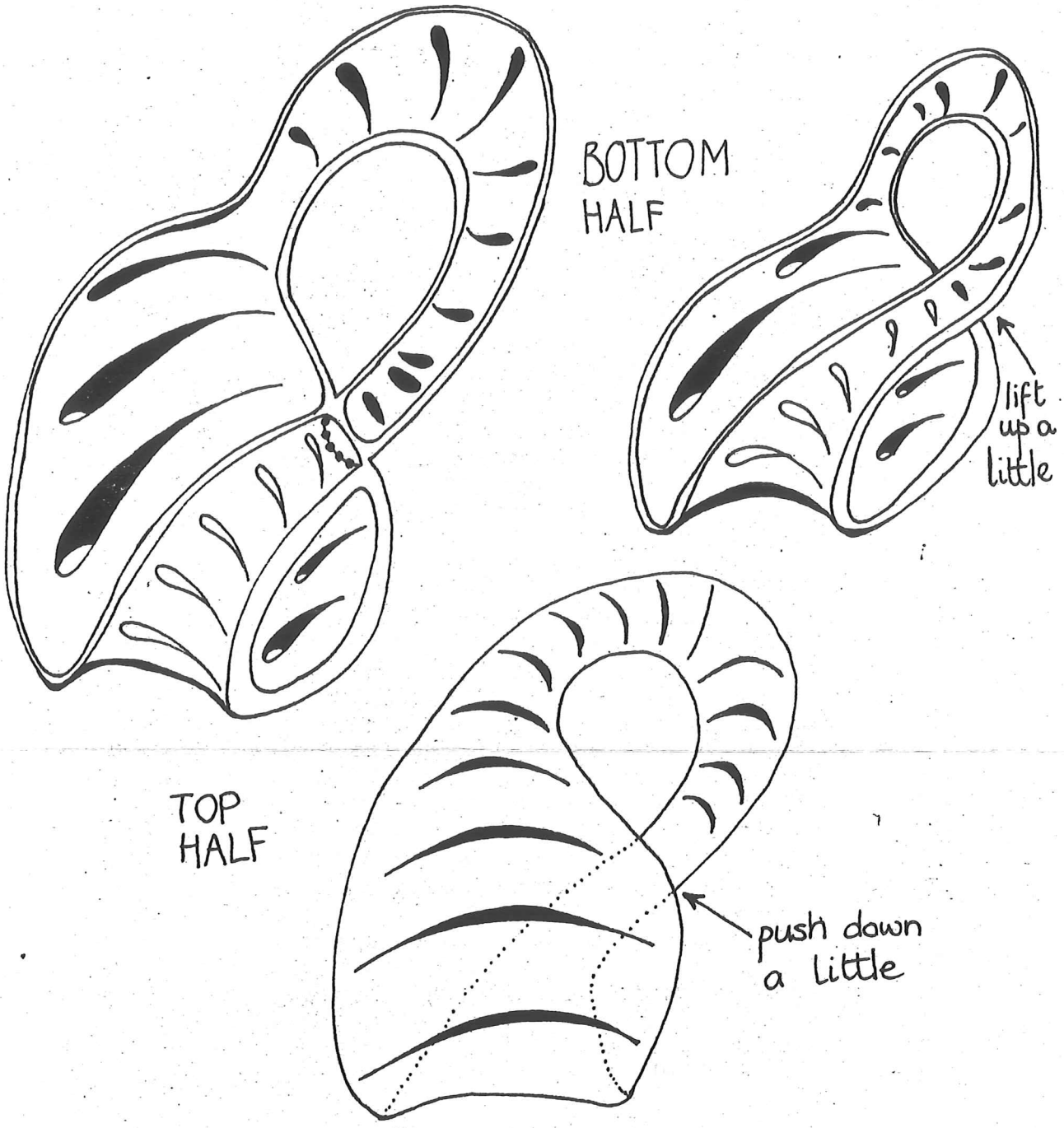
(ii) closed

(iii) triangulable.

Definition of orientability.

We call a surface orientable if it does not contain a Möbius strip; we call it non-orientable if it does contain a Möbius strip.

Of the ten examples above the first nine are all orientable, and only the last one, the Klein bottle, is non-orientable. To see that the Klein bottle contains a Möbius strip, imagine it to be sliced into two by the plane of the paper. In particular the self-intersection-circle C is sliced into two semicircles. Now slide the two pieces apart. Each piece will intersect itself in a semicircle, but we can remove these self-intersections by small moves as follows. Lift the thin part of the bottom piece up above the paper until it is clear of the rest; similarly push the thin part of the top piece down below the paper.

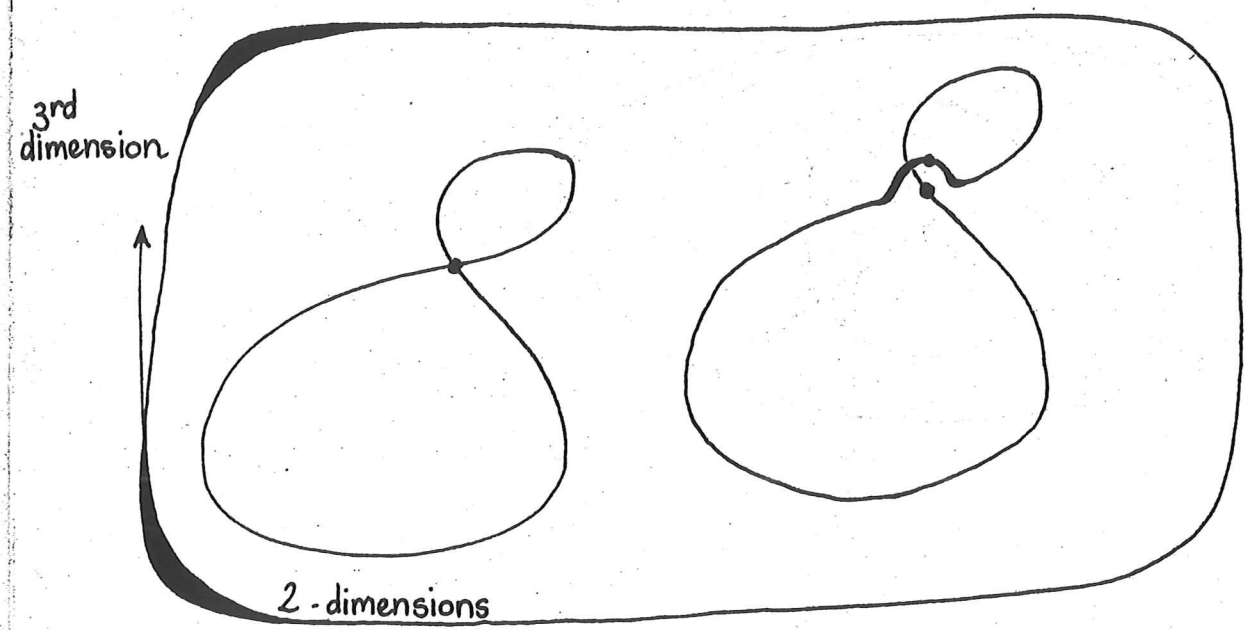


It can be seen that each piece is a Möbius strip. Therefore the Klein bottle not only contains a Möbius strip, but is in fact the the union of two Möbius strips sewn together along their boundaries.

Digression on one-sided surfaces.

It is true that the Möbius strip is one-sided. Some writers also call the Klein bottle one-sided, and claim that it has "no inside", but these statements are not true because of the self-intersection circle C. An ant cannot crawl from one side to the other because it would get held up at C; nor can it crawl from the inside to the outside. It is a theorem that the Klein bottle cannot be constructed in 3-dimensions without self-intersections and so this difficulty is fundamental. Moreover the same difficulty arises for any closed non-orientable surface. (The Möbius strip is not closed.) On the other hand it is possible to construct a Klein bottle in 4-dimensions without self-intersections: the way to do this is to lift the thinner tube a little way into the 4^{th} dimension, and then the self-intersection vanishes. At the same time the concept of the Klein bottle having an "inside" becomes meaningless, as the following analogy shows.

Consider a curve in 2-dimensions with one self-intersection point, like a figure 8. We can get rid of the self-intersection by lifting one branch a little way into the 3^{rd} dimension.



When we lift the figure 8 into the 3rd dimension it is meaningless to ask what becomes of the inside of the figure 8, because curves in 3-dimensions cannot have "insides" or "outsides". In exactly the same way surfaces in 4-dimensions cannot have insides or outsides, and therefore when we get rid of the self-intersections of a Klein bottle by lifting it into the 4th dimension, then it is meaningless to talk about it being one-sided. Therefore a Klein bottle cannot truthfully be said to be one-sided in either 3 or 4-dimensions. Consequently we prefer the term "non-orientable" to the term "one-sided". This ends the digression.

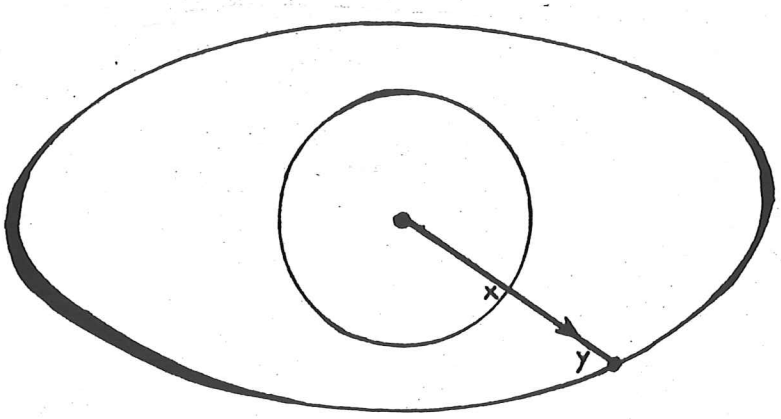
Definition of homeomorphism.

We now come to the central idea that distinguishes topology from any other form of geometry. Two surfaces X and Y are said to be homeomorphic if there is a one-to-one continuous function between them. We often write this as

$$X \cong Y \qquad \text{or} \qquad X \xrightarrow{\cong} Y.$$

We give some examples of homeomorphisms.

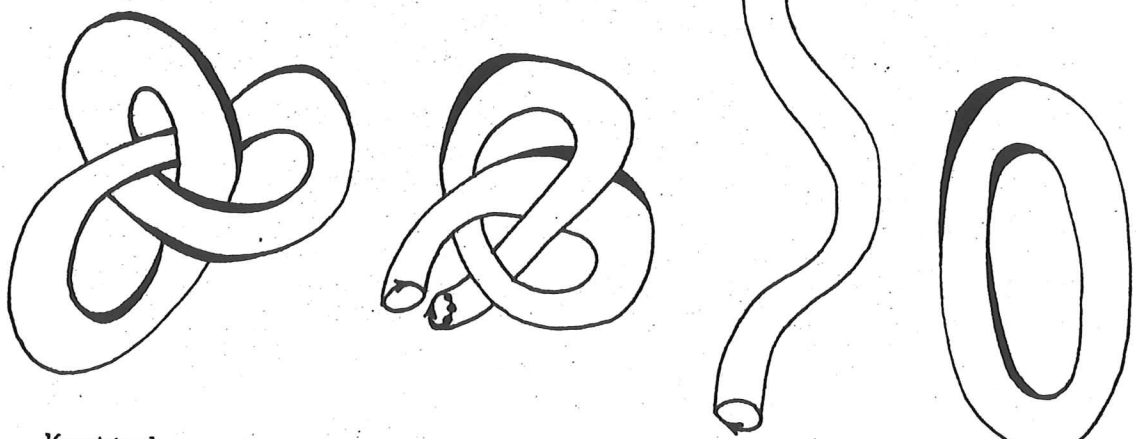
Example (i) A sphere is homeomorphic to an ellipsoid by radial projection $x \to y$.



Example (ii) More drastically, imagine the sphere X to be a rubber balloon, and bend into any shape Y (without cutting or glueing). Then $X \cong Y$, because each point of X is moved into a unique point of Y, and this determines a function $X \to Y$, which is continuous (since there is no cutting) and one-to-one (since there is no glueing). This example illustrates why topology is sometimes called rubber sheet geometry.

Example (iii) A sphere is homeomorphic to the surface of a tetrahedron. This example is important because it illustrates the fact that if we triangulate a surface X with curved triangles and make a model Y with straight triangles, then X is homeomorphic to Y.

Example (iv) Suppose that during a deformation of a surface X we made a cut, and later sewed the cut up again exactly as it was before, then the result would be homeomorphic to X. The following four pictures show, for example, that a knotted torus is homeomorphic to an unknotted torus.



1. Knotted torus
2. Cut (the arrows show the direction of cut)
3. Unknot the cylinder
4. Sew up again.

Digression on the knotted torus.

There is such a flavour of cheating about the last example that it is worth taking time off to explain why. Let K denote the knotted torus, let T denote the unknotted torus, and let E denote the 3-dimensional Euclidean space in which they are both embedded. We have two facts.

(1) There exists a homeomorphism $K \cong T$, and in Example (iv) above we explained how to construct such a homeomorphism.

(2) There is no homeomorphism of E onto itself throwing K onto T. Another way of saying this is that the knottedness of K is not a property of K by itself, but of the way it is embedded in E.

In this paper we study just the surfaces by themselves, and do not tackle the harder problem of how many ways they can be knotted in E (a problem which is still unsolved). Consequently to have them embedded in E is sometimes confusing, and sometimes raises red herrings like the knotting of tori, and the self-intersection and one-sidedness of the Klein bottle. We ought really to think of a surface as an abstract object existing on its own, without being embedded in anything. However this abstraction is a difficult concept for the beginner, and only becomes rigorous after familiarity with the foundations of analytic topology. Therefore we shall not insist upon it, and although the abstract concept will be implicit in our proofs, we shall continue to base our intuition on surfaces.

in E.

Exercises.

Prove that, of the ten examples at the beginning,

Example 2 $\cong$ Example 3 $\cong$ Example 4

Example 5 $\cong$ Example 6 $\cong$ Example 7

Example 8 $\cong$ Example 9.

What classification means.

We classify surfaces by inventing a list of standard surfaces and proving that every surface is homeomorphic to one of the standard ones. A more sophisticated way of saying this is that homeomorphism is an equivalence relation on the set of all surfaces, and we list the equivalence classes. Like many results in topology the classification theorem has a remarkable simplicity for the following reason. Homeomorphic surfaces can be so drastically different, that the equivalence classes are huge, and so there are very few of them, and the list is easy to compile.

Digression on the difference between topology and geometry.

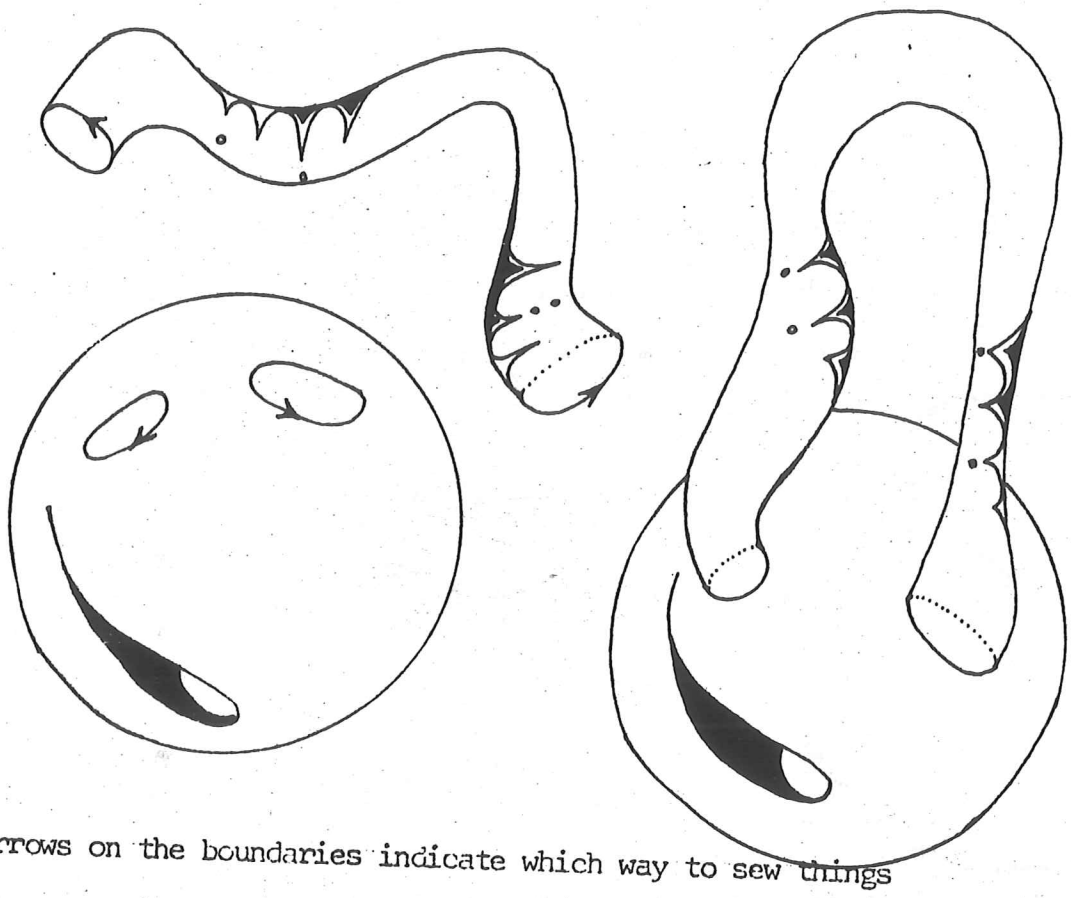
By comparison Euclidean geometry is much more complicated, because two surfaces are equivalent in Euclidean geometry if and only if one can be moved into the other by a rigid motion. Therefore the number of different surfaces, from the point of view of Euclidean geometry, is so enormous that nobody has even contemplated listing them. In other words topology can handle more complicated situations

than geometry, and yet with greater simplicity, because topology highlights the more dramatic properties of the situation and forgets the rest. For example from the geometrical point of view a torus has an inner diameter and an outer diameter, and many other measurements, but from the topological point of view the most dramatic property of the torus is its hole and the fact that it persists in having a hole however much we bend it about.

In 4-dimensions, and higher, the difference between geometry and topology becomes even more marked. Geometry becomes almost entirely algebraic in order to handle the complexity, while topology becomes more geometric in order to handle the simplicity. By "geometric" we mean that pictures are important both for furnishing the intuition to make conjectures, and for providing the inspiration to discover the proofs. The pictures can be on blackboards or in our imaginations, but wherever they are, it is true to say that topology is now the most geometric subject in mathematics.

The standard orientable surface of genus n.

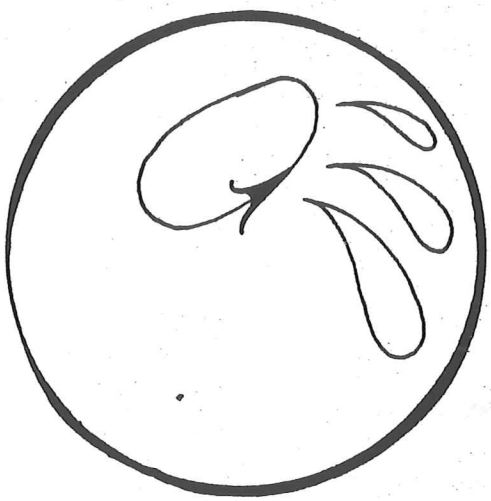
To sew a handle on a sphere, punch two little holes in the sphere, take a cylinder and sew the ends of the cylinder onto the boundaries of the holes.

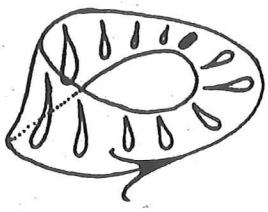


The arrows on the boundaries indicate which way to sew things together; notice that the two arrows on the cylinder go the same way, but that the two arrows on the boundaries of the holes go opposite ways. A sphere with a handle sewn on is homeomorphic to a torus (if we had got one of the arrows reversed it would have been a Klein bottle). If we want to sew a number of handles on we sew them on to different parts of the sphere. Examples 6 and 9 at the beginning show spheres with 2 and 3 handles sewn on. Define <u>the standard orientable surface of genus n</u> ($n \geqslant 0$) to be a sphere with n handles sewn on. In particular genus 0 means a sphere, genus 1 a torus, and genus 2 a pretzel.

The standard non-orientable surface of genus n.

To sew a Möbius strip on a sphere, punch one little hole in the sphere, take a Möbius strip and sew the boundary of the Möbius strip onto the boundary of the hole. This sounds simple, and from the abstract point of view it is as simple as it sounds, but if we





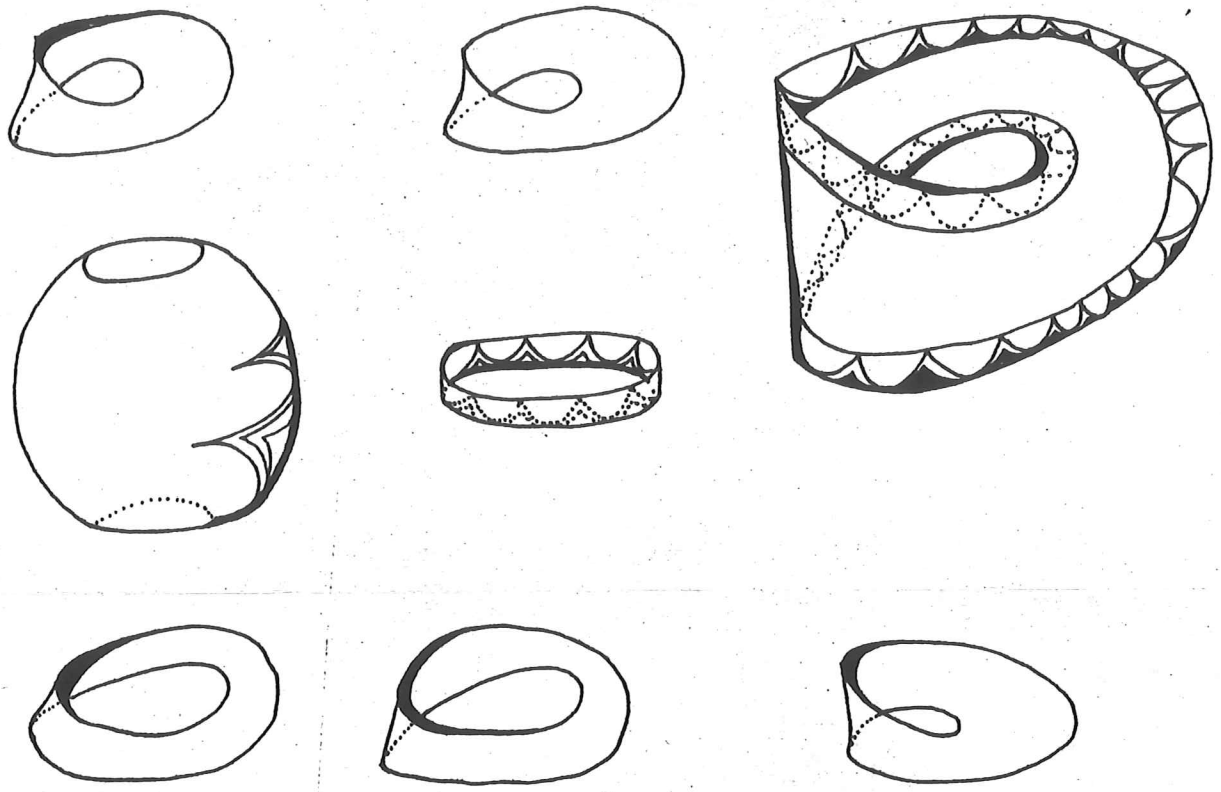
want to visualise it happening in 3-dimensions then it is not obvious, because the resulting surface has to intersect itself. In a moment we shall give an alternative description that is easier to visualise. Meanwhile define the standard non-orientable surface of genus n ($n \geq 1$) to be a sphere with n Möbius strips sewn on.

Examples.

(i) The case n = 0 is omitted because this would give a sphere, which is orientable.

(ii) The case n = 1 gives a surface called the real projective plane, which is described further in Appendix 2.

(iii) The case $n = 2$ gives the Klein bottle because sewing two Möbius strips on a sphere is homeomorphic to sewing two Möbius strips together, as the following pictures show.



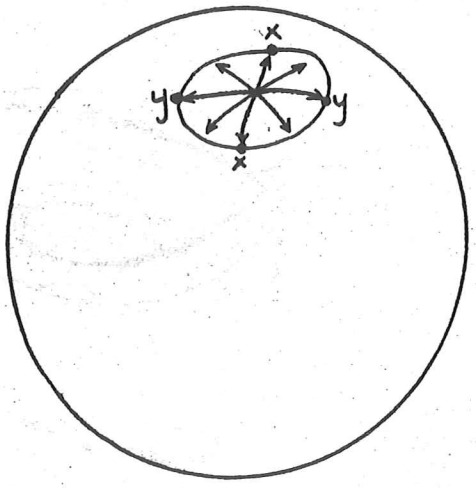
1. Sphere with two Möbius strips sewn on.
2. Shrink the sphere part into a little cylinder.
3. Sew the cylinder onto one of the Möbius strips.

Alternative descriptions of sewing on Möbius strips.

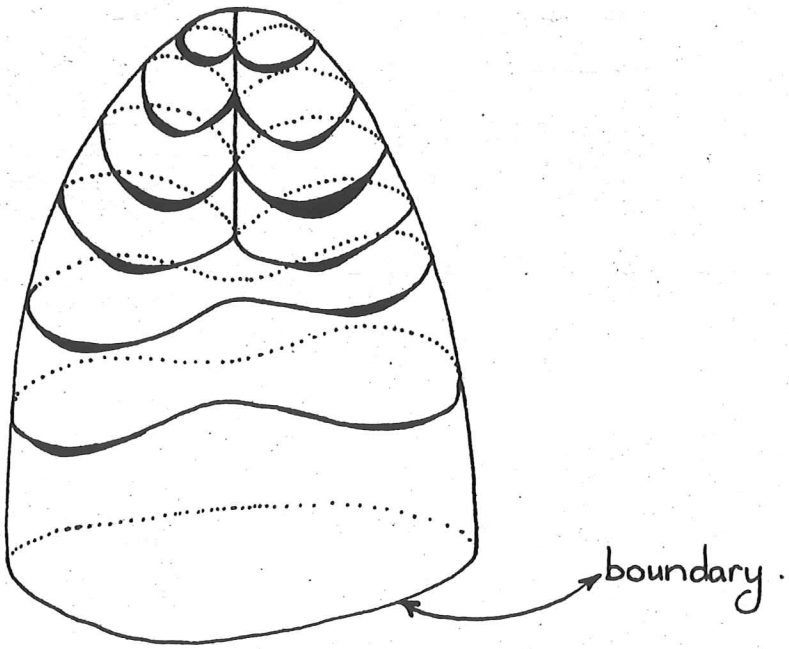
We shall show that the following three processes are equivalent.

(1) Sew a Möbius strip on a sphere.

(2) Punch a hole in the sphere and then sew together all pairs of diametrically opposite points on the boundary of the hole. We abbreviate this by saying "sew diametrically".

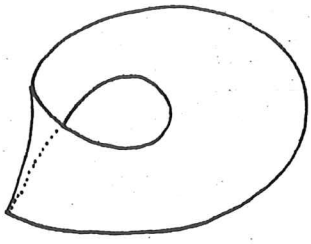


(3) Punch a hole in the sphere and then sew on a cross-cap, where a cross-cap is the surface illustrated in the following picture.

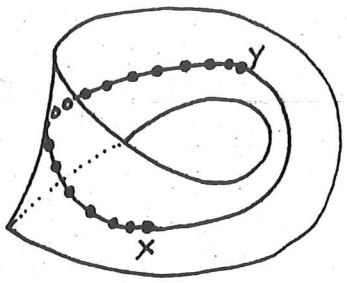


The cross-cap has a boundary, and intersects itself in the line xy. The horizontal cross-sections are drawn in to indicate how it intersects itself. We shall show that a cross-cap is the same as a Möbius strip with self-intersections.

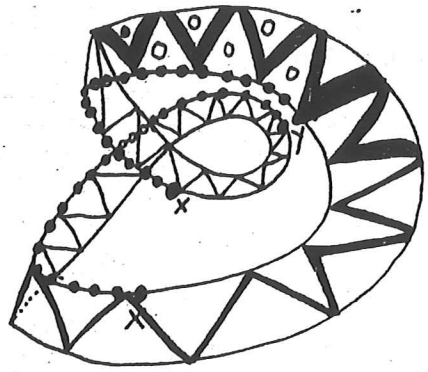
To see that (1) is equivalent to (2) consider the following sequence of pictures. Starting with (1), cut the Möbius strip along its centre line, and it is a well-known party trick that it does not fall apart, but becomes a twisted cylinder, which is homeomorphic to an untwisted cylinder. Sew the base of the cylinder onto the sphere, and there remains to sew up the top again diametrically, which is homeomorphic to prescription (2).



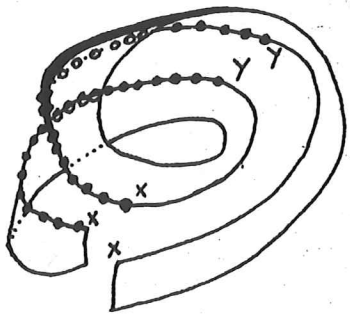
1. Möbius strip.



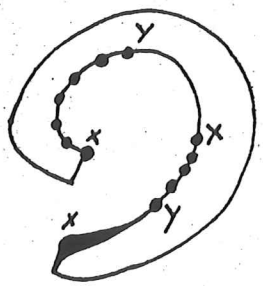
2. Draw in centre line.



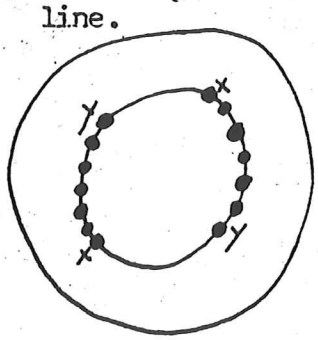
3. Cut along centre line.



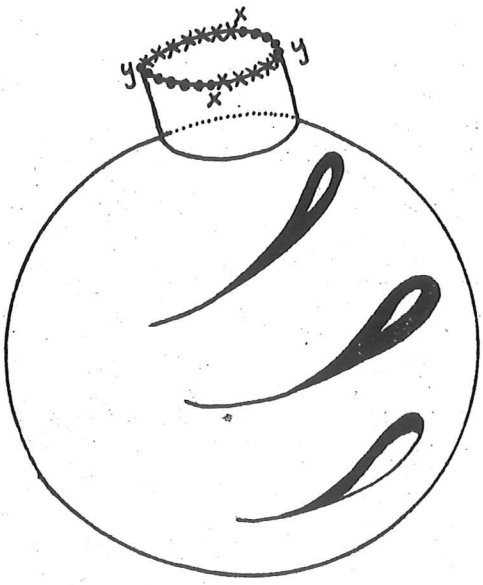
4. Cut across.



5. Untwist.

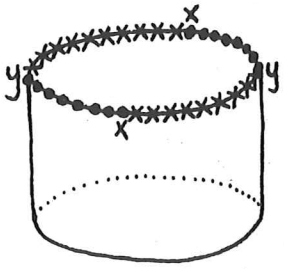


6. Sew up the last cut.

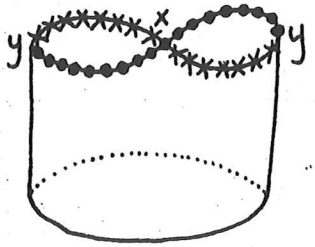


7. Sew onto the sphere.

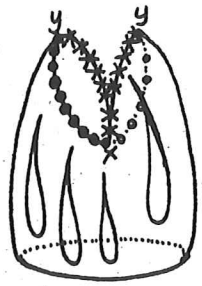
To see that (2) is equivalent to (3) consider the following sequence of pictures. Sew the top of the cylinder diametrically to form a cross-cap, by first pinching the ends of one diameter x, then the ends of another diameter y, then sewing together one pair of arcs xy, and finally the other pair of arcs yx.



1. Take the cylinder.

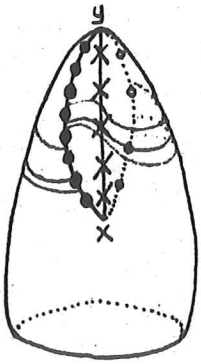


2. Pinch x.

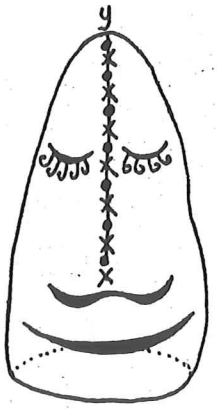


3. Fold up.

4. Pinch y.



5. Sew up xy.



6. Finally sew up yx to form cross-cap.

We conclude by deducing that the standard non-orientable surface of genus n cross-caps on a sphere. This is easier to visualise than sewing on Möbius strips, but is aesthetically less pleasing from the abstract point of view, because of the self-intersections.

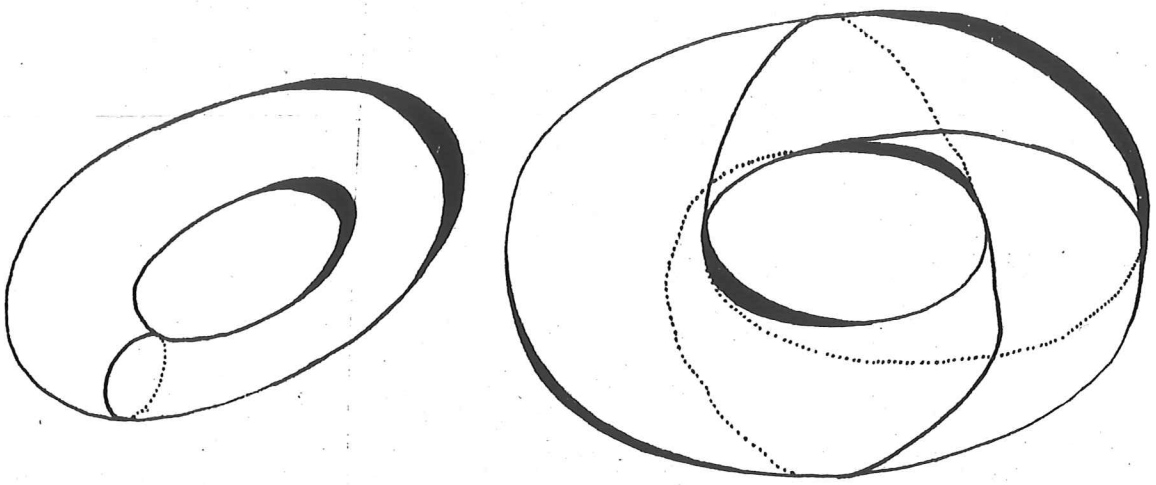
We are now in a position to state the main theorem. The statement will resemble a watershed dividing the first half of the paper from the second, because up till now we have been developing topological intuition so that the reader can fully understand the meaning of the theorem, whereas from now on we shall be concentrating on technique in order to prove the theorem.

Classification Theorem. Any connected closed triangulable surface is homeomorphic to one of the standard ones.

Before proving the theorem we need some definitions and lemmas (a lemma is a little subsidiary theorem).

Definition of spherelike.

Let M be a surface. Choose a triangulation of M. By a curve on M we mean a closed path without self-intersections, consisting of vertices and edges of the triangulation. The reason that we use the word "curve" for something that sounds more like a polygon is that it conjures up the correct intuition. Whenever we talk about a curve on a surface, it is easier to think of it without the triangulation. Here are two examples of curves on a torus.



A curve is said to separate M if cutting along the curve causes M to fall into two pieces. We call M spherelike if every curve (in every triangulation) separates M.

Example (i). The sphere is spherelike - this is the famous Jordan Curve Theorem, which we shall prove in Lemma 2. It is also the justification for using the word spherelike.

Example (ii). The torus is not spherelike, because neither of the two curves shown above separate it.

Definition of the Euler characteristic.

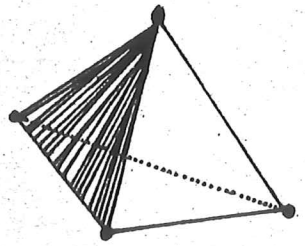
Let M be a surface. Choose a triangulation of M. Let v be the number of vertices, e the number of edges and t the number of triangles. The Euler characteristic $\chi(M)$ is defined by the formula

$$\chi(M) = v - e + t.$$

The remarkable fact about $\chi(M)$ is that it is independent of the triangulation, although of course v, e and t depend upon the triangulation. The result for spheres was first published by Euler in 1752, although was probably known to Archimedes in the second century B.C. The result for other surfaces was discovered by Poincaré in the 1890's, and out of this small germ grew the whole of algebraic topology. The first rigorous proof of the invariance of $\chi(M)$ was not until the 1930's, and is beyond the scope of this paper (see Appendix 1). The formula works not only for triangulation with triangles, but also for "triangulations" with polygonal faces.

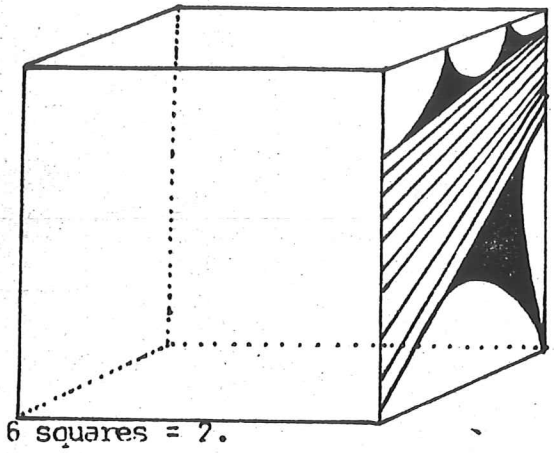
Examples

1. Tetrahedron.



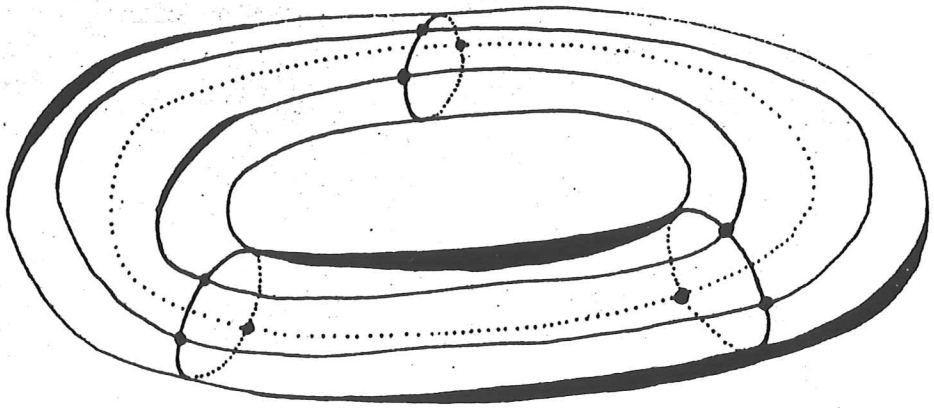
χ = 4 vertices - 6 edges + 4 triangles = 2.

2. Cube.



χ = 8 vertices - 12 edges + 6 squares = 2.

3. Torus



χ = 9 vertices - 18 edges + 9 squares = 0

Exercises.

By triangulating suitably show that

1. χ(cylinder) = 0.

2. χ(Mobius strip) = 0.

3. χ(sphere with n holes punched in it) = 2 - n.

Deduce that

4. χ(standard orientable surface of genus n) = 2 - 2n.

5. χ(standard non-orientable surface of genus n) = 2 - n.

We now state two lemmas. The programme will be first to prove the theorem using the lemmas, and then to prove the lemmas (with the help of further lemmas). The reason for doing it this way round is to give motivation for the lemmas.

Lemma 1. If M is a connected closed triangulable surface then $\chi(M) \leq 2$.

Lemma 2. If M is a connected closed triangulable surface then the following three conditions are equivalent:

(a) M is sphere like

(b) $\chi(M) = 2$

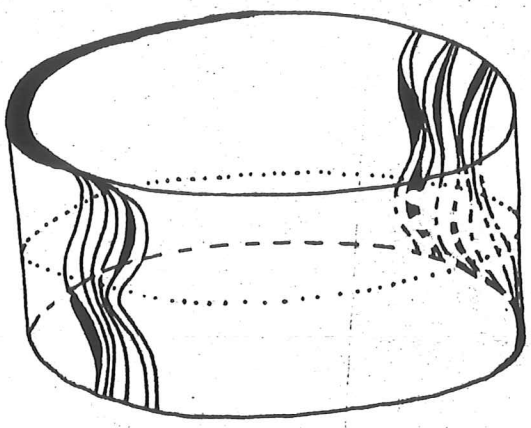
(c) M is homeomorphic to a sphere.

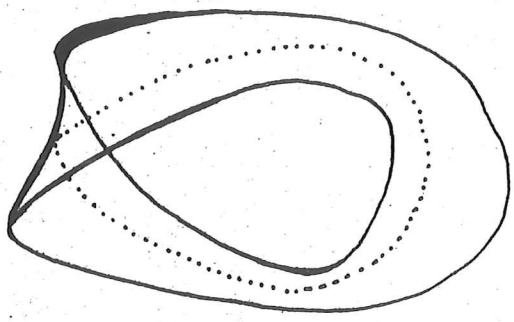
Proof of the Theorem.

Let M be a given connected closed triangulable surface. We have to prove that M is homeomorphic to one of the standard ones.

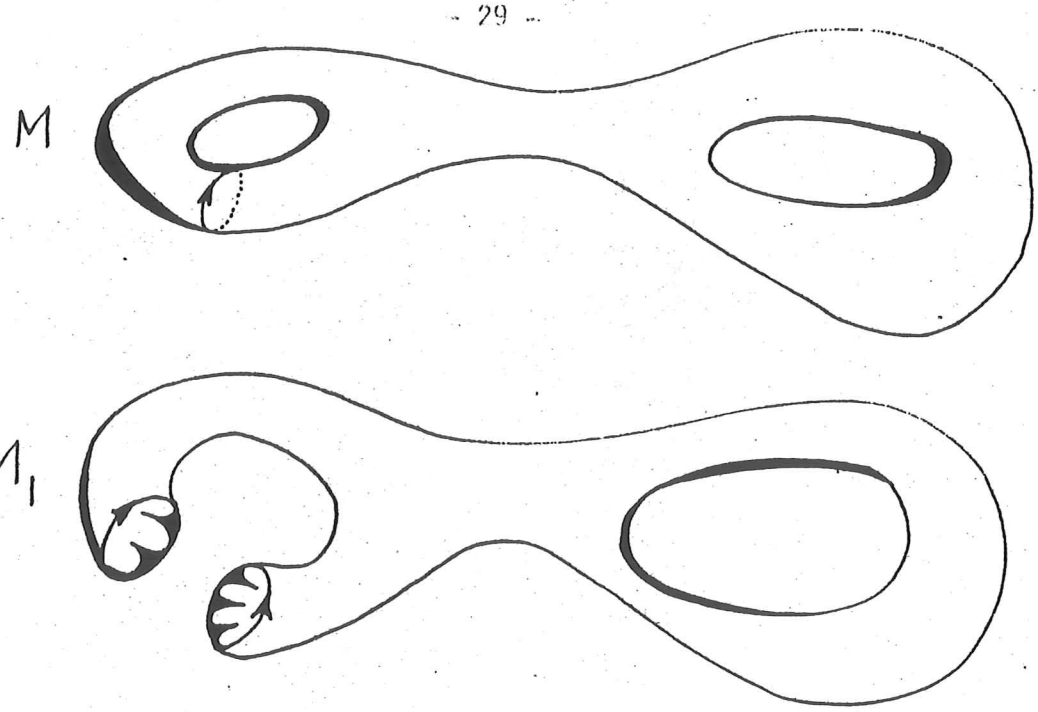
Choose a triangulation of M, and compute $\chi(M)$. Then $\chi(M) \leq 2$ by Lemma 1. If $\chi(M) = 2$ then M is homeomorphic to a sphere by Lemma 2. Therefore assume $\chi(M) < 2$. Therefore M is not spherelike by Lemma 2, and so we can choose a curve C not separating M.

Consider a thin strip of surface containing C. There are two possibilities: the strip is either a cylinder or a Möbius strip.





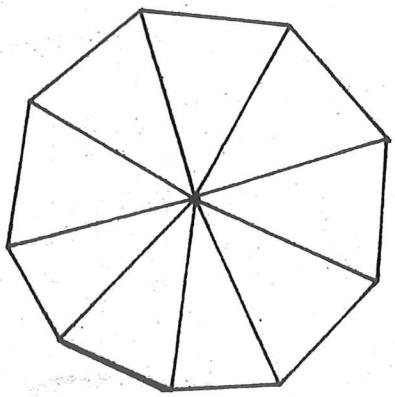
If it is a cylinder we call C an orientation-preserving curve on M, and if it is a Möbius strip we call C orientation-reversing. We now construct a new surface M_1 by a process called surgery, which is defined as follows. If C is orientation-preserving, cut along C and fill in each side with a disk. It is important to leave the arrows on the boundaries of the disks in order to remind us which way to sew them up again later on.



If C is orientation-reversing, again cut along C, but this time only one boundary curve is formed instead of two, so fill this in with one disk instead of two. If the surgery is performed on an abstract surface there is no difficulty about self-intersections, but if the surgery is performed in 3-dimensions then new self-intersections may arise. We claim that

$$\chi(M_1) = \begin{cases} \chi(M) + 2, \text{ if C is orientation-preserving} \\ \chi(M) + 1, \text{ if C is orientation-reversing.} \end{cases}$$

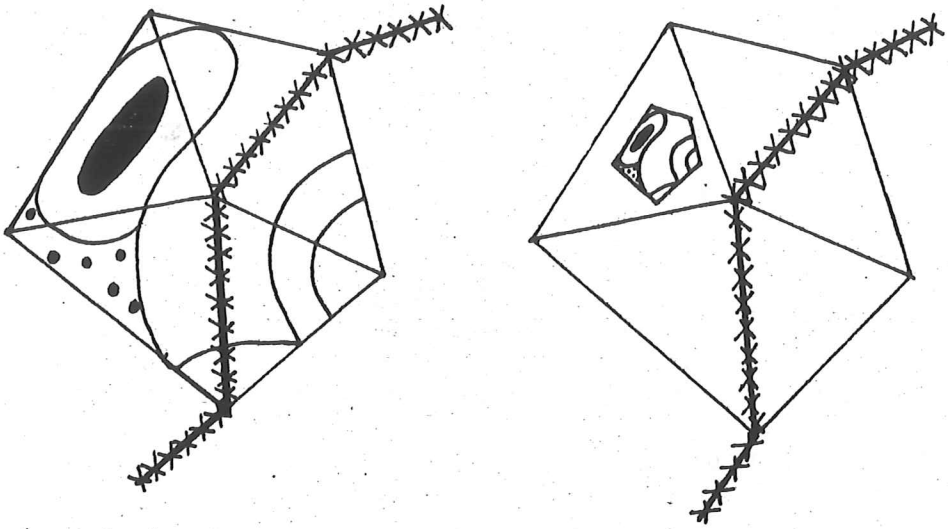
To prove this suppose that C contains k vertices and k edges. Then $\chi(C) = k - k = 0$. Therefore removing C does not alter $\chi(M)$. In the orientation-preserving case we form M_1 by adding two disks, where each disk is obtained by joining C to a point.



Therefore each disk contains (k+1) vertices, 2k edges and k triangles. Therefore $\chi(\text{disk}) = 1$ and $\chi(M_1) = \chi(M) + 2\chi(\text{disk}) = \chi(M) + 2$. In the orientation-reversing case only one disk is added; although this disk contains 2k triangles it still has characteristic 1, and so $\chi(M_1) = \chi(M) + 1$. In both cases $\chi(M) < \chi(M_1)$.

We now proceed inductively. Either $\chi(M_1) = 2$ and M_1 is homeomorphic to a sphere, or else $\chi(M_1) < 2$ and we can surger M_1 into M_2 where $\chi(M_1) < \chi(M_2)$. By Lemma 1 the process must stop after a finite number of steps, and so we obtain a finite sequence of surfaces $M, M_1, M_2, \ldots, M_r$ such that $\chi(M) < \chi(M_1) < \chi(M_2) < \ldots < \chi(M_r) = 2$, with M_r homeomorphic to a sphere by Lemma 2. As an exercise the reader is recommended to draw the pictures for the examples at the beginning of the paper.

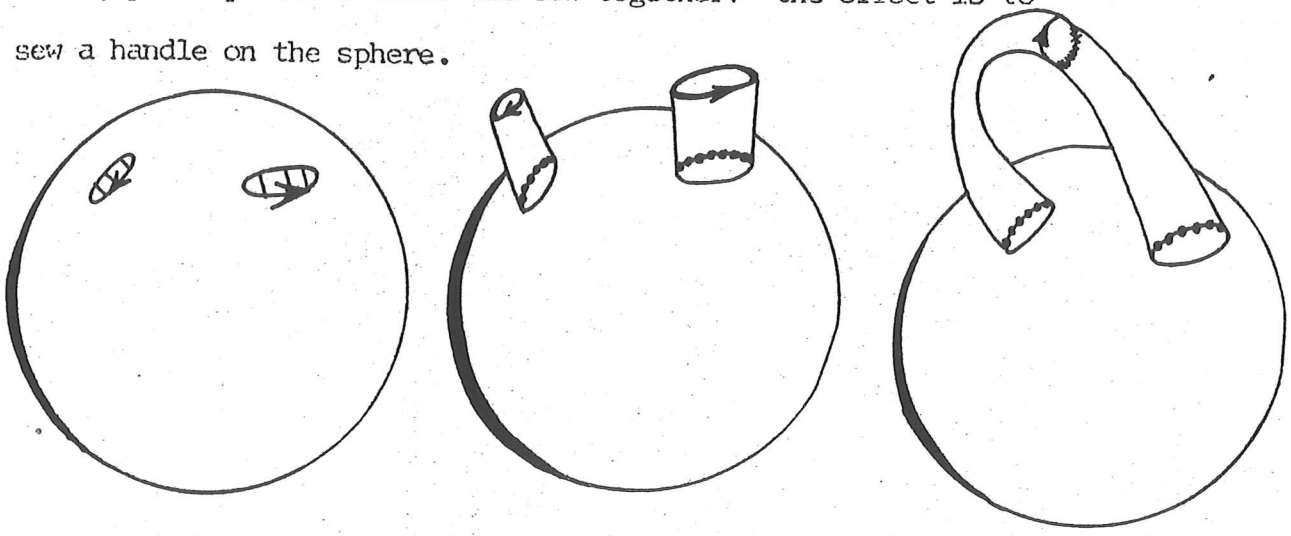
Now M_r contains a number of little disks arising from the surgeries, and we can ensure that all these disks are disjoint by the following trick. The only way in which two disks might not be disjoint would be if the curve of a later surgery cut across the disk of an earlier surgery. The trick is to shrink each disk into the interior of one of its triangles, because this will ensure that it automatically misses any later curves.



The triangulation has served its purpose, and we now forget it; we concentrate only on the disks in M_r. Imagining M_r to be made of rubber, we can move together each pair of disks that arose from an orientation-preserving surgery. More precisely, we can choose the homeomorphism from M_r to the sphere, so as to bring each pair close together.

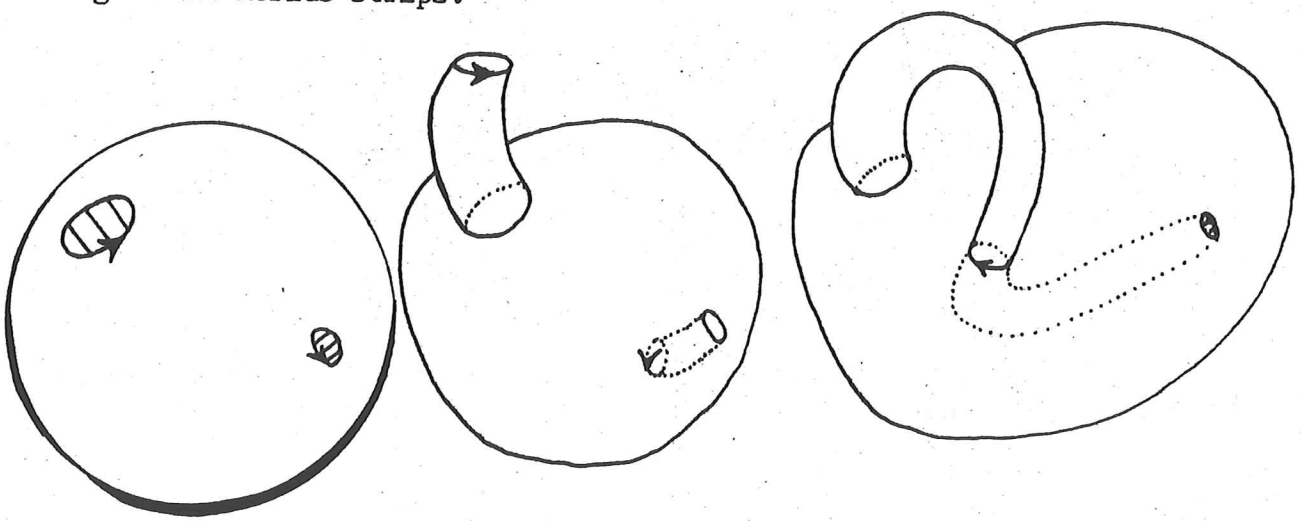
We then desurger, as follows. There are three types of desurgery.

Type (i) Two disks with arrows going opposite ways. Remove the disks, push up little tubes and sew together: the effect is to sew a handle on the sphere.



Type (ii) One disk. Remove the disk and sew the boundary diametrically: the effect is to sew on a Möbius strip.

Type (iii) Two disks with arrows going the same way. Remove the disks, push one tube up and one tube down, bend round and sew together. The effect is to sew on a Klein bottle, which is equivalent to sewing on two Möbius strips.



Performing all the desurgeries simultaneously we obtain a surface M_* homeomorphic to the original M.

The orientable case.

If M is orientable so is M_*. Therefore M_* contains no Möbius strips, and so only desurgeries of type (i) can occur. Therefore M_* is a standard orientable surface, namely a sphere with n handles sewn on. The genus, n, is the number of surgeries (or desurgeries) and therefore can be computed from the Euler characteristic

$$n = 1 - \tfrac{1}{2}\chi(M).$$

The non-orientable case.

If M is non-orientable then all three types of desurgery can occur. If only type (i) occurs, then M_* will be orientable, which is a contradiction, and therefore at least some of types (ii) and (iii) must occur. Perform these first. We now use a trick to convert all type (i) desurgeries into type (iii), as follows. Given a pair of disks corresponding to a type (i) desurgery, transport (i.e. pull along rubber-wise) one of the disks round the sphere to one of the Möbius strips that has already been sewn on, round the Möbius strip, and back. Like the ant, which found itself on the other side after crawling round a Möbius strip, so the arrow on the disk will be going the other way round after its transportation. In other words the type (i) will be converted into type (iii). Therefore M_* is a standard non-orientable surface, namely a sphere with n Möbius strips sewn on. The genus, n, is the

number of type (ii) desurgeries plus twice the number of type (iii), and therefore can be computed

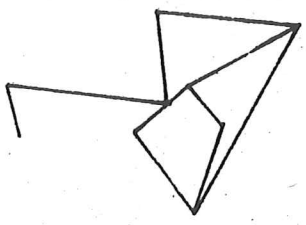
$$n = 2 - \chi(M).$$

This completes the proof of the Theorem.

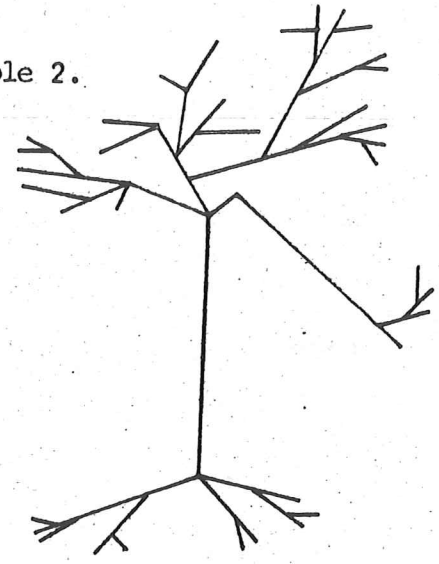
Graphs.

In order to prove Lemmas 1 and 2 it is necessary to introduce graphs. A graph is a connected set of vertices and edges. Connected means that it is all in one piece, or equivalently that any two vertices are connected by a path in the graph.

Example 1.



Example 2.



There are two possibilities: a graph may or may not contain loops. Example 1 contains loops, but Example 2 does not. A graph that contains no loops is called a tree.

Lemma 3. A tree always contains at least one end vertex (i.e. a vertex on only one edge).

Proof. Suppose not. Suppose that every vertex lies on two or more edges. Then, starting at any vertex, it is possible to proceed along a path in the graph, such that each edge is followed by a different edge. If we continue for more steps than there are vertices, then we must have performed a loop, which is a contradiction. Therefore the lemma is true.

If G is a graph with v vertices and e edges, then the Euler characteristic $\chi(G) = v - e$.

Lemma 4. If T is a tree then $\chi(T) = 1$.

Proof. The proof is by induction on the number e of edges in T. The induction begins with $e = 0$, for then T is a point and so $\chi(T) = 1 - 0 = 1$. Suppose the lemma true for $e - 1$, and suppose we are given a tree T with e edges. By Lemma 3 choose an end vertex. Removing that vertex and the edge containing it will not alter $\chi(T)$, and will leave a tree, T_1 say, with $e - 1$ edges. Therefore $\chi(T) = \chi(T_1) = 1$, by induction.

Lemma 5. If G is a graph containing a loop then $\chi(G) < 1$.

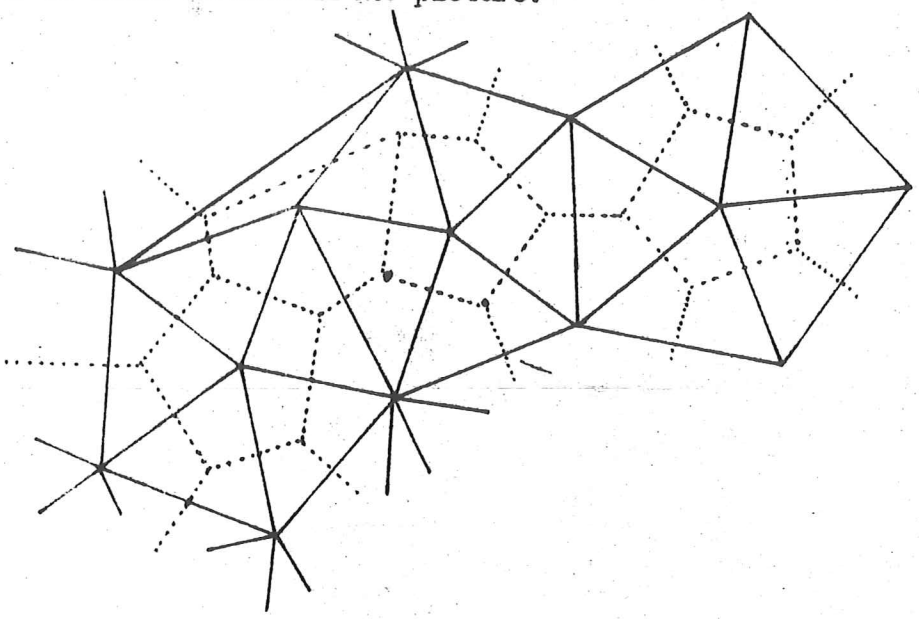
Proof. Since G contains a loop, we can remove one edge from that loop without disconnecting G, and therefore obtain a graph, G_1 say, such that $\chi(G) = \chi(G_1) - 1$. Either G_1 is a tree, or else we can remove another edge to form a graph G_2. We can go on removing edges until we hit a tree, and we must hit a tree after a finite number of steps because e is finite. Suppose, therefore, that G_r is a tree, $r \geq 1$.

Then

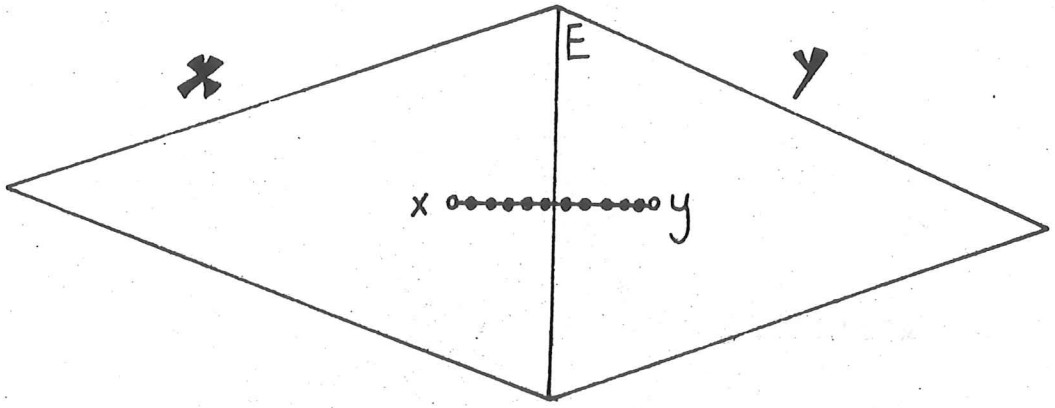
$$\begin{aligned}\chi(C) &= \chi(G_r) - r \\ &= 1 - r, \text{ by Lemma 4} \\ &< 1.\end{aligned}$$

Dual-triangulations.

Let M be a connected closed triangulable surface. Choose a triangulation of M. The dual-triangulation is defined as follows. It is shown dotted in the picture.



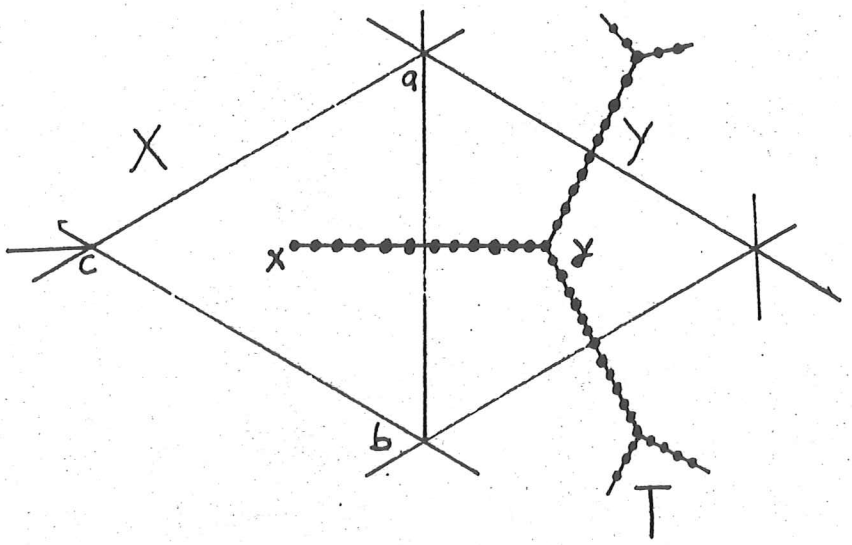
Within each triangle X choose an interior point x, and call it the dual-vertex of X. If two triangles X, Y have an edge E in common, join their dual-vertices x, y to form a dual-edge xy. The dual-edge xy intersects E once, and does not meet any other edges.



Any tree consisting of dual-vertices and dual-edges is called a dual-tree. The complement K of a dual-tree T is defined to be the set of all vertices, edges and triangles of M that do not meet T.

Lemma 6. The complement K of a dual-tree T is connected.

Proof. Since K contains all the vertices of M, it is enough to prove that any two vertices of M can be connected by a path along the edges of K. The proof is by induction on n, the number of edges in T. The induction begins with n = 0, for then T consists of one dual-vertex, so K contains all the edges of M, and therefore K is connected because M is connected. Now assume the result true for n - 1. Given a dual-tree T with n edges, choose an end dual-vertex x by Lemma 3, and let xy be the dual-edge of T containing x. Let X, Y be the triangles with dual-vertices x, y and let the vertices of X be a, b, c as shown.

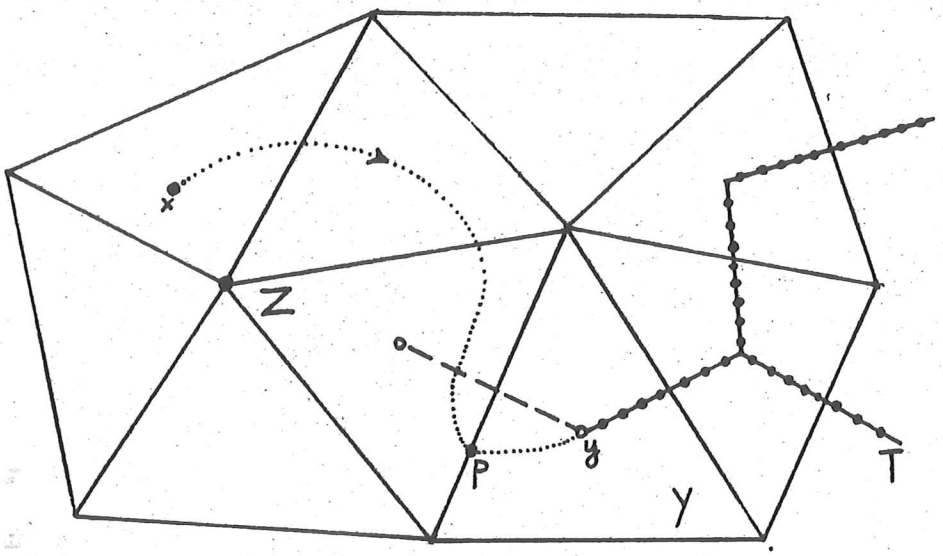


Let T_1 be the dual-tree obtained from T by removing x and xy, and let K_1 be the complement of T_1. Then K_1 is connected by induction. But K is obtained from K_1 by removing the triangle X and the edge ab. This does not disconnect K_1 because any path in K_1 containing the edge ab can be replaced by a path in K containing edges ac and cb. Therefore K is connected.

Lemma 7. A maximal dual-tree contains all the dual-vertices.

Proof. Let T be a maximal dual-tree, that is to say T is not contained in any larger dual-tree. Suppose that T does not contain the dual-vertex x. Then we shall prove a contradiction.

For, let P be a path from x to any point of T. By shifting P slightly we can make sure that it does not go through any vertices.



Let p be the first point on the path P that lies in a triangle Y, whose dual-vertex y lies in T. Since p is the first such point, it must lie on some edge of Y (and not at a vertex by construction). Let Z be the other triangle containing this edge. Then the dual-vertex z of Z does not lie in T (otherwise p would not have been the first such point). Let T_1 be the dual-tree obtained by adding yz and z to T. Therefore T_1 is larger thar T and so T is not maximal, and the contradiction proves Lemma 7.

Proof of Lemma 1.

Let M be a connected closed triangulable surface, and choose a triangulation of M. Let T be a maximal dual-tree, and let G be the

complement of T. Then T contains all the dual-vertices by Lemma 7, and so G contains no triangles. Therefore G consists of vertices and edges, and is connected by Lemma 6. Therefore G is a graph. There are one-to-one correspondences between

vertices of M $\longleftrightarrow$ vertices of G

edges of M $\longleftrightarrow$ edges of T and G

triangles of M $\longleftrightarrow$ vertices of T.

Therefore $\chi(M) = \chi(T) + \chi(G)$

$= 1 + \chi(G)$, by Lemma 4

≤ 2, by Lemma 5.

Proof of Lemma 2.

We have to show that the following three statements are equivalent: (1) M is spherelike

(2) $\chi(M) = 2$

(3) M is homeomorphic to a sphere.

We shall prove that (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (1). First we prove that (1) implies (2). Therefore assume M is spherelike, and suppose $\chi(M) \neq 2$, and we shall prove a contradiction. Let T be a maximal dual-tree and let G be its complementary graph. Then

$$\chi(G) = \chi(M) - \chi(T) = \chi(M) - 1 \neq 1.$$

Therefore G is not a tree, and consequently contains a loop C.

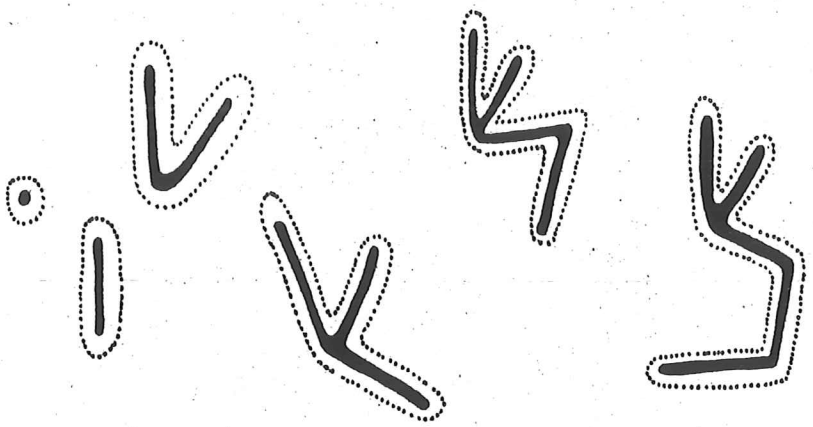
This loop is a curve* on M, and therefore separates M into two pieces, because M is spherelike. Each piece of M contains at least one triangle, and therefore at least one dual-vertex. But all dual-vertices are contained in T by Lemma 7, and any two can be connected by a path in T because T is a tree, and a tree is connected. This path does not meet G, the complement of T, and therefore does not meet the curve C. In other words points in the two pieces of M can be joined by a path not meeting C. Therefore C does not separate M after all, which is a contradiction. This completes the proof that (1) implies (2).

We now prove that (2) implies (3). Therefore assume that $\chi(M) = 2$. We have to show that M is homeomorphic to a sphere. Let T be a maximal dual-tree, and let G be the complementary graph. Then G is also a tree by Lemma 5, because $\chi(G) = \chi(M) - \chi(T) = 1$.

---

* The reader may ask why we use two different words "loop" and "curve" for apparently the same thing. The reason is that each gives the correct intuition of its context: a loop in a graph is like a loop in an electronic network, while a curve is something that one draws on a surface.

Let N(T) be a neighbourhood of T formed by thickening T. We claim that N(T) is homeomorphic to a disk, and the proof is as follows. By applying Lemma 3 inductively shrink T to a point by retracting edge after edge. Starting from the point we can reverse the process by expanding out edge after edge. Now put a little disk around the point, and every time an edge expands out grow out an arm amoeba-like to contain this edge. At the end the disk has grown homeomorphically into N(T).

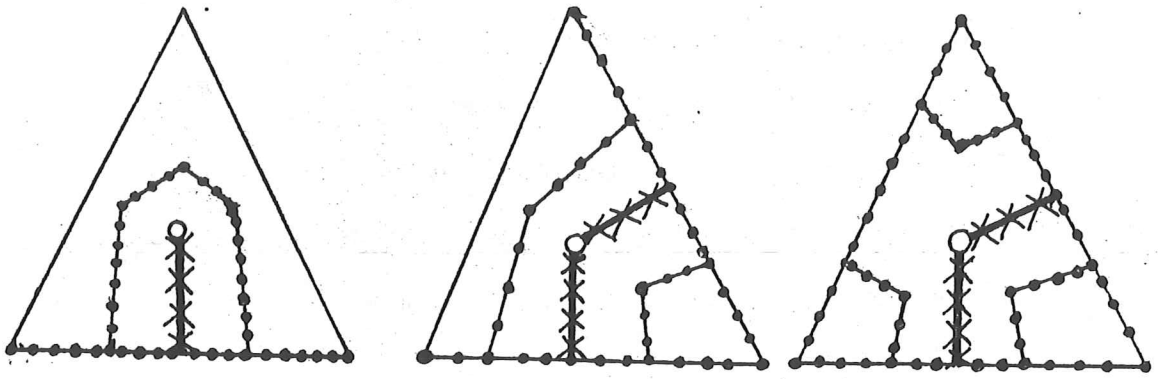


Similarly a neighbourhood N(G) of the tree G is homeomorphic to a disk. The idea is to choose N(T) and N(G) so that their union is the whole surface M, and their intersection is the boundary of each. This can be done as follows. First make a model of the surface in which the edges are straight, the triangles are flat, and the intersections of triangles and dual-edges are straight. Given a point x in the model, let X denote the triangle containing x (or one of the triangles containing x, if x happens to lie on an

edge), and let $t(x)$, $g(x)$ denote the distances from

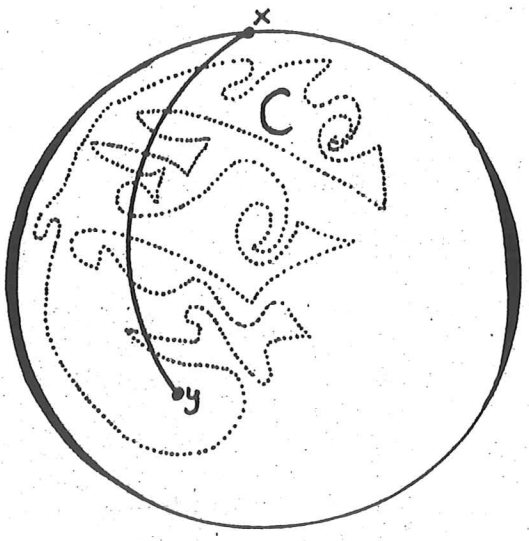
$$x \text{ to } T \cap X,\ G \cap X$$

respectively. Put x into $N(T)$ if $t(x) \leq g(x)$, and into $N(G)$ if $g(x) \leq t(x)$. The intersection $N(T) \cap N(G)$ consists of points such that $t(x) = g(x)$, and is the boundary of both. The intersections of $N(T)$ and $N(G)$ with typical triangles are shown **below.**

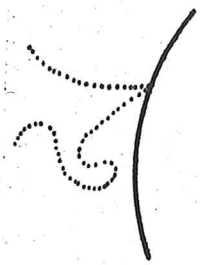


Consequently $M = N(T) \cup N(G)$ is homeomorphic to two disks sewn along their boundaries, namely a sphere. This completes the proof that (2) implies (3).

Finally we prove that (3) implies (1), that a sphere is spherelike. In other words we have to show that any curve C on a sphere separates the sphere. This is a polygonal form of the celebrated Jordan Curve Theorem. Assume that C is a polygon consisting of a finite number of great-circle arcs.



Choose a point x on the sphere, not on C, nor on any of the great-circles containing arcs of C. Regard x as the north pole. Given any other point y, not on C, nor at the south pole, define y to be <u>even or odd</u> according as to whether the number of intersections of C with the meridional arc xy is even or odd. We have a convention that an intersection like



counts as 0 or 2. With this convention we deduce that all points near an even point are even, and all points near an odd point are odd. Therefore along any path not crossing C the parity remains

constant. In other words no even point can be joined to an odd point without crossing C, and so C separates the sphere into evens and odds. The poles give no trouble because x is even, and if the south pole is not on C it has unambiguous parity. This completes the proof of Lemma 2, and hence also completes the proof of the Classification Theorem.

We conclude the paper with four appendices.

1. For the experts.
2. The real projective plane.
3. Why non-orientable surfaces have to have self-intersections.
4. Some problems on knots.

Appendix 1. For the experts.

The discerning reader will have observed that we managed to get through the paper without ever defining a surface. And deliberately so. For the topologist, the correct definition is as follows: a surface is a 2-dimensional locally-Euclidean compact connected Hausdorff space. The advantages of this definition are that it is intrinsic, it is in topological terms, and it generalises immediately to higher dimensions. But we did not introduce it at the beginning because it is too technical: it cannot be understood without reading a book on analytic topology (for example Reference 5). We wanted to present geometry to the beginner, and it is tough going for the beginner to have to first plough through the foundations of analytic topology.

The second reason for omitting the definition was to avoid the problem of triangulability. The only correct published proof that a surface can be triangulated is Reference 6, which is much too hard for the beginner. In assuming triangulability we implicitly defined a surface to be a collection of triangles fitting together according to the rules (i) each edge lies on two triangles, and (ii) each vertex is joined to a polygon. With this implicit definition our proofs were rigorous. However it would perhaps have been a little unaesthetic to make this into an explicit definition, because it sounds rather artificial. And indeed we should have been in danger of confusing the concept with the tool. The concept of surface is one of man's richest intuitions, whereas a triangulation is a mere mathematical tool. Therefore our taste was to present the ideas of "intuition and tool" and avoid any "definition" that was too technical or artificial.

The other major omission was the proof of the topological invariance of χ. Invariance is necessary in order to prove that two surfaces of different genera are not homeomorphic. The proof is hard and requires the full power of algebraic topology (see for example Reference 4). One expresses χ in terms of homology groups, and then proves the homology groups to be topological invariants.

It is true in Lemma 2 we proved that $\chi(\text{sphere}) = 2$, but

the result depended upon our proof of the Jordan Curve Theorem,

$$\text{sphere} \Longrightarrow \text{spherelike}$$

which was only for triangulations made up of great-circle arcs. To get a topologically invariant proof the triangulation must be allowed to be arbitrarily wiggly. The proof that we gave breaks down because an arbitrary curve may meet a meridian in an infinite number of points. We remark, however, that this part of our proof was not really necessary, and only put in for good measure. The careful reader will have noticed that the only parts of Lemma 2 that were used in the proof of the classification theorem were the other two parts:

$$\text{spherelike} \Longrightarrow \chi = 2 \Longrightarrow \text{sphere}.$$

We conclude the appendix with some remarks about current research. We claimed at the beginning that the classification theorem was a beautiful example of geometric topology, and this is because it has strong overtones of higher dimensions. The definition of an n-manifold is an n-dimensional locally-Euclidean compact connected Hausdorff space. In other words a surface is the same as a 2-manifold (this was why we used the letter M to denote a surface). The classification of 2-manifolds (surfaces) was achieved in the last century, and for the whole of this century topologists have been struggling in vain to classify 3-manifolds. The main stumbling block is the celebrated Poincaré Conjecture, which is the 3-dimensional analogue of

$$\text{spherelike} \Longrightarrow \text{sphere}.$$

Poincaré conjectured this in 1899, but it is still unsolved today. In spite of this, the analogous Poincaré Conjecture in dimensions ≥ 5 was solved in 1961 by Smale and others. During the last five years there have been spectacular successes in high dimensional geometric topology, which have given us new insight into the low dimensions. The new proof of the classification theorem that we have given above is an example of this insight.

Appendix 2. The real projective plane.

We begin by emphasising the word "real", because often it is not clear whether a writer is discussing real projective geometry or complex projective geometry. There are three definitions of the real projective plane, P.

(1) P is the set of lines through the origin in Euclidean 3-dimensions.

(2) The projective plane is $\pi \cup \pi_\infty$ where π is the Euclidean plane, and π_∞ the set of "points at infinity".

(3) The standard non-orientable surface of genus 1, namely a disk and a Möbius strip sewn together.

We shall show that all three definitions are equivalent. From the geometrical point of view definition (1) is the most elegant, because it also contains the linear structure of P. Definition (2) is aesthetically bad because finite points are different from infinite points, whereas all points P are qualitatively the same. Definition (3) is the topologists viewpoint.

To show (1) implies (2), let π be a plane in 3-dimensions not through the origin. Then each line through the origin meets π in a unique point, or else is parallel to π, in which case we say that it meets π in a "point at infinity". Therefore there is a one-to-one correspondence $P \longleftrightarrow \pi \cup \pi_\infty$.

To show (1) is equivalent to (3), let S be the unit sphere centre the origin. Each line through the origin meets S in a pair of antipodes. Therefore there is a one-to-one correspondence

points of P $\longleftrightarrow$ pairs of antipodes of S.

Therefore to recover P, sew S together antipodally. First sew the northern hemisphere onto the southern hemisphere, and then sew the equator diametrically. This is equivalent to sewing a Möbius strip on a sphere.

Appendix 3. Why non-orientable surfaces have to have self-intersections.

Theorem. Any closed connected surface in 3-dimensions has an inside and an outside.

Proof. The proof is analogous to that of the Jordan Curve Theorem. Let M be the given surface in 3-dimensions. Choose a point x well away from M. Given any other point y not on M, call y even or odd according as to whether the number of intersections of xy with M is even or odd. Define x to be even. Then the even points form the

outside, and the odd points the inside, and there is no path from an inside point to an outside point without crossing M.

Corollary. Any closed non-orientable surface in 3-dimensions must have self-intersections.

Proof. Suppose not. Suppose we have M without self-intersections. Without loss of generality we can assume M is connected. By the theorem M has an inside and an outside, and since M is non-orientable it contains a Möbius strip. If we start an ant on the inside, and let it crawl round the Möbius strip, then it will finish up on the outside, and will have traced a path from inside to outside without crossing M, which is a contradiction.

Appendix 4. Some problems on knots.

The following problems are harder than they look, but playing with them may help the reader to sharpen his intuition. A general introduction to knot theory is Reference 3.

1. The torus in Example 3 at the beginning of the paper is knotted on the outside because the outside is not homeomorphic to the outside of an unknotted torus. Similarly the torus in Example 4 is knotted on the inside. Prove that a torus cannot be knotted on both the inside and the outside at the same time, (see Reference 1).

2. Prove that any curve in 4-dimensions can be unknotted.

3. Prove that a sphere can be knotted in 4-dimensions (see Reference 2).

4. Prove that any sphere in 5-dimensions can be unknotted (see Reference 7).

5. Prove that two spheres can be linked in 5-dimensions.

REFERENCES

1. J.W. Alexander : On the subdivision of a 3-space by a polyhedron, Proc. Nat. Acad. Sci. 10 (1924), 6.
2. E. Artin : Zur Isotopic zweidimensionalem Flächen in R_4, Abh. Math. Sem. Hamburg 4 (1926), 174-177.
3. R.H. Crowell and R.H. Fox : Introduction to knot theory, (Ginn) 1963.
4. P.J. Hilton and S. Wylie : Homology theory (Cambridge) 1960.
5. B. Mendelson : Introduction to topology (Blackie) 1963.
6. T. Radó : Über den Begriff der Riemannschen Fläche, Acta. Litt. Sci. Szeged, 2 (1925), 101-121.
7. E.C. Zeeman : Unknotting spheres in five dimensions, Bull. Amer. Math. Soc. 66 (1960), 198.

Mathematics Institute
University of Warwick
Coventry

June 1966