THE WORK OF JOHN MILNOR

W.T. GOWERS

1. The hardest IQ question ever

What is the next term in the following sequence: 1, 2, 5, 14, 41, 122? One can imagine such a question appearing on an IQ test. And one doesn't have to stare at it for too long to see that each term is obtained by multiplying the previous term by 3 and subtracting 1. Therefore, the next term is 365.

If you managed that, then you might find the following question more challenging. What is the next term in the sequence

$$1, 1, 28, 2, 8, 6, 992, 1, 3, 2, 16256, 2, 16, 16, 523264?$$

I will leave that question hanging for now, but I promise to reveal the answer later.

2. A giant of modern mathematics

There are many mathematicians with extraordinary achievements to their names, and many who have won major prizes. But even in this illustrious company, John Milnor stands out. It is not just that he has proved several famous theorems: it is also that the areas in which he has made fundamental contributions have been very varied, and that he is renowned as a quite exceptionally gifted expositor. As a result, his influence can be felt all over modern mathematics. I will not be able to do more than scratch the surface of what he has done, partly because of the limited time I have and partly because I work in a different field.

The areas that he has worked in include differential topology, K-theory, group theory, game theory, and dynamical systems. He has written several books that have become legendary for their high quality. Princeton University Press describes one of his best-known books thus:

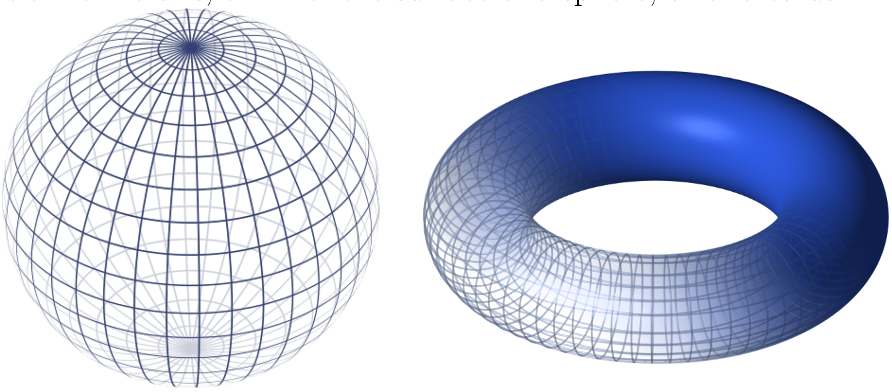
> One of the most cited books in mathematics, John Milnor's exposition of Morse theory has been the most important book on the subject for more than forty years.

I recently read the following sentiment on the website Mathoverflow: "Usually when Milnor explains, things are easy ..." One reviewer on Amazon describes Milnor's book, Topology from the Differentiable Viewpoint as the "best math book ever written". These are just opinions, of course, but similar opinions are held by many people.

Milnor was awarded a Fields medal in 1962, the Wolf Prize in 1989, and is the only person to have won all three AMS Steele Prizes (for seminal contribution to research in 1982, for mathematical exposition in 2004, and for lifetime achievement this year). And now he has been awarded the Abel Prize. Let me attempt to explain why.

3. Differential structures on spheres

One of the most important concepts in mathematics is that of a *manifold*. To get an idea of what a manifold is, think of the surface of a sphere, or of a torus.

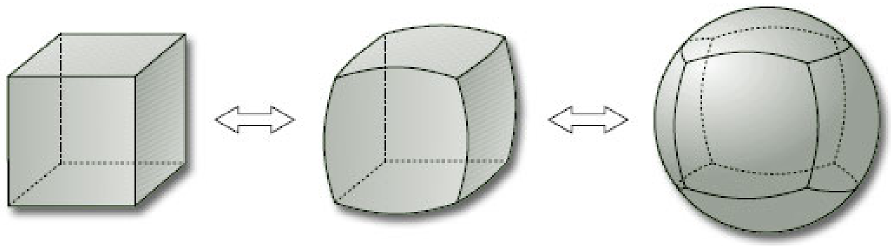


If you cut out a very small piece of one of these surfaces, then it looks like a flat sheet of paper (though in fact it won't necessarily be *perfectly* flat). A sheet of paper is two-dimensional, so we say that these surfaces are two-dimensional as well. We also call them "locally Euclidean". This means that if you look at a very small region in such a surface, then its geometry is just like the ordinary two-dimensional geometry of the plane, the geometry we are all familiar with that goes back to Euclid. However, and this is what makes manifolds interesting, the *global* behaviour of manifolds is not Euclidean at all. For instance, if you take two parallel lines on the surface of a sphere – where "line" no longer means straight line but does still mean the shortest path between two points – and continue them for long enough, then they will eventually meet. (For example, imagine two lines that point due north from the equator. They will start out parallel and will meet at the North Pole. This can be seen in the picture above.)

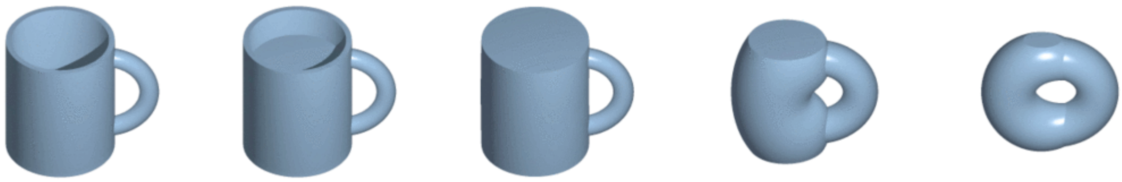
There are three further important points to make about manifolds. The first is that although the examples I have just given, a sphere and a torus, are most naturally visualized as surfaces that live inside three-dimensional space, it is possible to talk about manifolds *intrinsically*. That is, one can discuss the geometry of a manifold by mentioning just the points in the manifold itself and making no reference to any external space inside which the manifold lives. Secondly, there can be manifolds of any dimension. It is not easy to visualize 23-dimensional manifolds, but as abstract concepts they exist every bit as much as the more familiar two-dimensional surfaces.

The third point is essential background to one of John Milnor's most remarkable results. In my description of manifolds above, I said that the main characteristic feature of a manifold is that its geometry is locally just like Euclidean geometry. I deliberately left the words "just like" a little vague, because it turns out that there are several different notions of "just like" that can be useful here. I would like to focus on just two of these.

The first notion comes from topology: we say that two shapes are *homeomorphic* if each one is a continuous deformation of the other. For example, the surface of a cube is homeomorphic to the surface of a sphere,



and (to give a famous example) the surface of a teacup is homeomorphic to the surface of a doughnut.



If we have a shape of which we can say that all small enough patches are homeomorphic to a little patch of Euclidean space, then we call that shape a *topological manifold*.

The second notion is more demanding and comes from calculus: we say that two shapes are *diffeomorphic* if they are not just *continuous* deformations of each other but *differentiable* deformations of each other.

What does this mean? Well, very roughly speaking, a continuous deformation is one where you are allowed to stretch your shape as much as you like, but are not allowed to tear it. A differentiable deformation is "smoother" than a continuous one is required to be. It has the additional property that if you follow a steady path in the first shape, then the corresponding path in the deformed shape should not have any sudden changes of speed or direction. Folds, corners, sharp bends: these are disallowed.

In 1956, Milnor found an extraordinary mathematical object: a shape that is homeomorphic to a seven-dimensional sphere but not diffeomorphic to a seven-dimensional sphere. He called this object an "exotic sphere". I would like to spend a little bit of time explaining why exotic spheres are so counterintuitive, but once I have done that there is a danger that I will have persuaded you that they cannot exist at all, so I will then have to try to explain how in fact they can.

First, let me express in a different way what it means for exotic spheres to exist. It means that you can take a sphere and continuously deform it to form another shape that is also smooth, in the sense of not having any "corners"; however, even though the original sphere is smooth and the continuously deformed sphere is smooth, the only *deformations* that turn one into the other are *not* smooth.

The reason this is so surprising is that one's intuitions suggest that it ought to be possible to "iron out the kinks" in any continuous deformation between the two shapes. For example, suppose we take a one-dimensional sphere – otherwise known as a circle. (I call this one-dimensional because I am just talking about the curve and not its interior.) In this case, a smooth shape homeomorphic to the original will just be a *smooth closed curve* – that is, a curve that starts and ends in the same place and does not cross itself or have any sharp bends. Now it is perfectly possible to find continuous deformations from the circle to the distorted circle that are not smooth: as you go round the original circle, a corresponding point will go round the distorted circle, and it might well have sudden changes of speed. However, it is intuitively clear (and can be proved mathematically) that if there are sudden changes of speed, they are completely unnecessary: instead of instantaneous jumps in speed we could have rapid accelerations, and would end up approximating our continuous deformation by a smooth deformation.

It is very tempting to think that one could do something like this process not just in one dimension but in any number of dimensions. The idea would be this: given a continuous deformation between a sphere and some distorted, but still smooth-looking, sphere, we would look at the places where the continuous deformation failed to be smooth and would modify the deformation at those places to make it smooth.

Why does this not work? One of the reasons this is hard to explain is that the impossibility of carrying out this smoothing process is a *global* phenomenon rather than a *local* one. That is, if you look at a small patch of the original sphere and the corresponding small patch of the distorted sphere, you *can* iron out any kinks that there might be. However, in a high number of dimensions, there is a much bigger variety of possible kinks than one finds in just one dimension. For example, if the spheres are five-dimensional, then the places where the kinks occur could quite easily form a complicated three-dimensional set. The effect of this is that when you iron them out in one place, you may simply push the crease to somewhere else (an experience I sometimes have when ironing a shirt).

If you are lucky, you can push the crease until it meets itself and cancels out. But Milnor's example shows that this will not always happen. One sign of just how unexpected this was comes from Milnor's own reaction to the discovery:

> When I first came upon such an example in the mid-50's, I was very puzzled and didn't know what to make of it. At first, I thought I'd found a counterexample to the generalized Poincaré conjecture in dimension seven. But careful study showed that the manifold really was homeomorphic to S^7. Thus, there exists a differentiable structure on S^7 not diffeomorphic to the standard one.

The fact that manifolds could be homeomorphic but not diffeomorphic meant that the differentiable manifold was an important object in its own right, and not just a way of looking at a topological manifold. For this reason, Milnor's construction gave birth to a whole new field of mathematics, which is known as differential topology and which includes several other highlights of modern mathematics.

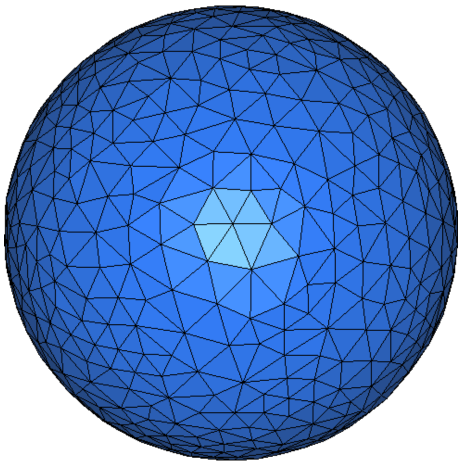
What do these "exotic spheres" actually look like? Another surprise is just how simple they are to define. This is true of Milnor's original construction, but it is slightly easier to describe a later construction that became known as a "twisted sphere". (In fact, Milnor's example can itself be viewed as a twisted sphere, even though he did not explicitly define it as one.) Just as a sphere is made from two parts that are topologically equivalent to discs – the northern and southern hemispheres – so a seven-dimensional sphere can be made out of two seven-dimensional discs. These discs are joined at a six-dimensional equator. Note that there are two copies of the equator – one at the boundary of the northern hemisphere and one at the boundary of the southern hemisphere – and we glue each point to the same point in the other copy. But what if we were to join points in the southern copy to *different* points in the northern copy, by stretching the two copies in some directions and squashing them in others? If we do this in the right way, it turns out that we obtain a manifold that is homeomorphic to a sphere but not diffeomorphic to a sphere: that is, an exotic sphere.

Let me now return to the mysterious sequence from the beginning of this article. After Milnor's discovery that there could be "exotic" differential structures on spheres, the question naturally arose of how many genuinely distinct such structures there were in each dimension. With Michel Kervaire, Milnor worked this out for several different dimensions: the sequence gives the answers, starting at dimension five. Thus, there is only one differential structure in five dimensions (the standard one) and six dimensions. In seven dimensions there are 28. The sequence takes us up to 19 dimensions, where there are 523264 differential structures. As it happens, the next term in the sequence is 24, so that is the answer to the IQ test question. These remarkable numbers are related to other topological phenomena and also to the seemingly very different field of number theory.

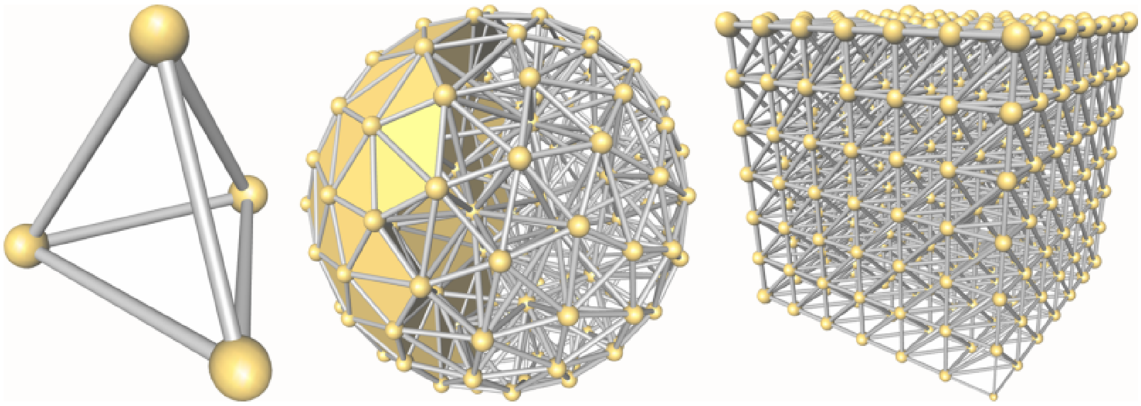
4. Further geometrical results

In this section, I shall discuss some other beautiful results of John Milnor. However, I shall be briefer about these.

4.1. **The Hauptvermutung.** One of the techniques we have for studying curved geometrical shapes such as manifolds is to *triangulate* them. Here, for example, is a triangulated sphere.

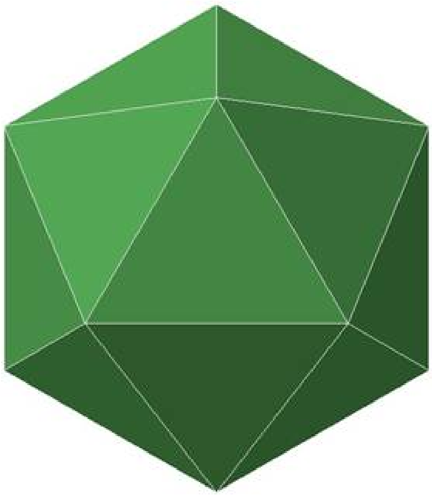


And here is a picture that gives some indication of what a triangulation of a region of three-dimensional space can look like.



Triangulating a manifold breaks it up into little pieces (triangles or tetrahedra or higher-dimensional versions of these) that are simple to understand. The hope then is that one can study the manifold by taking careful account of how these little building blocks are put together.

However, there are difficulties with this idea. It seems pretty obvious that every manifold can be triangulated (that is, continuously deformed into a manifold that is built out of "triangles" of the appropriate dimension). But how does one prove this? Also, the same manifold can be triangulated in many different ways. For example, we saw above a triangulation of a two-dimensional sphere. Here is a much simpler one.

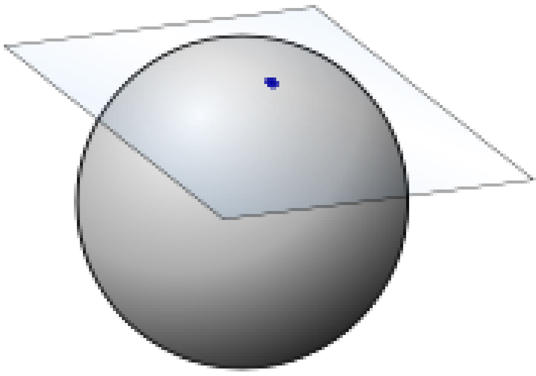


One would like to be confident, when studying a manifold with the help of a triangulation, that different triangulations give rise to the same conclusions about the manifold. A good way to do that, it turns out, is to show that the triangulations have a *common refinement*. To see what this means, suppose you have a triangulation and you divide up the triangles into smaller triangles. This gives a new triangulation, which we call a *refinement* of the original one. It is usually easy to show that if you refine a triangulation then the refined triangulation will have the same essential properties as the original one. This means that if you start with two different triangulations and find a new triangulation that refines *both* the original triangulations, then you know that the two triangulations have the same essential properties.

In two dimensions, triangulations always exist, and any two triangulations have a common refinement. That makes it tempting to believe that the same will be true in all dimensions. The *Hauptvermutung* (German for "main conjecture") asks whether any two triangulations have a common refinement.

Actually, there are two versions of this problem, one for manifolds and one for more general objects known as "triangulable space". In 1961 Milnor gave a counterexample to the more general version of the Hauptvermutung, which had been an open problem since 1908. A few years later Andrew Casson and Dennis Sullivan disproved the version for manifolds as well. And in 1982 Michael Freedman discovered a four-dimensional manifold that cannot be triangulated at all.

4.2. **The hairy ball theorem and parallelizable spheres.** At each point of a two-dimensional sphere, we can draw a *tangent plane*.

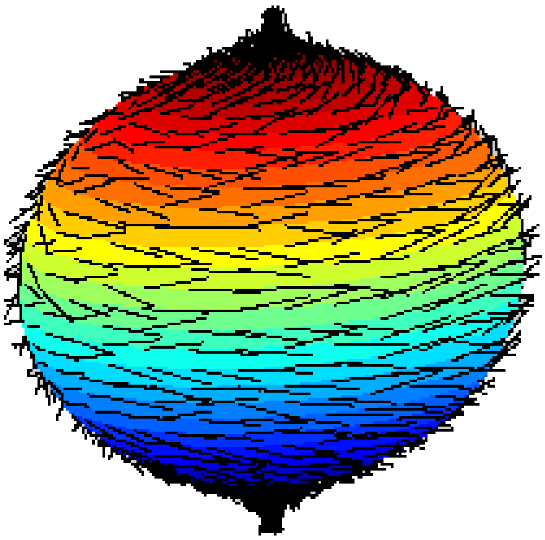


These tangent planes are of fundamental importance in the theory of differentiable manifolds, because they are needed if one wishes to make sense of the notion of the derivative of a map. Roughly speaking, we think of the tangent plane at a point as being the plane that best approximates the manifold near that point. For higher-dimensional manifolds, we talk of a tangent *space*, but the basic idea is the same.

One thing we like to do with any space is choose a coordinate system. In the two-dimensional case illustrated, that would mean that we would think of each point as the origin of its tangent space, and would decide how to draw an x-axis and a y-axis, and also which was the positive direction along each axis and which the negative direction. All this information can be encoded in two arrows, one pointing along the x-axis in the positive direction and one pointing along the y-axis in the positive direction.

A very natural question then arises: can we choose our arrows at the various points in such a way that they vary continuously when the points themselves vary continuously? If we can, then the sphere is called *parallelizable*.

In the two-dimensional case, one of the classic results of algebraic topology, proved by Brouwer in 1912, shows that we cannot. In fact, it is not even possible to choose *one* axis in a continuous way. This result is known as the hairy ball theorem, because it can be thought of as saying that there is no way of combing the hair of a hairy ball without having points where the hair bunches up.



However, for the three-dimensional sphere things are very different. We can think of this sphere as consisting of all points of the form (x, y, z, w) such that $x^2 + y^2 + z^2 + w^2 = 1$. Now each such point can be regarded as an object called a *quaternion*, a kind of "number" invented in 1843 by William Rowan Hamilton. Quaternions are a bit like complex numbers, but we now put in *three* square roots of -1, which are traditionally called i, j and k. An odd, but essential, feature of quaternions is that the order in which you multiply them matters: for example $ij = k$ but $ji = -k$. The quaternion associated with the point (x, y, z, w) is $x + yi + zj + wk$, and if $x^2 + y^2 + z^2 + w^2 = 1$ then it is called a *unit quaternion*. It turns out that if you multiply two unit quaternions together, then you get another unit quaternion. Also, if you take any unit quaternion q, then the three quaternions qi, qj and qk are all perpendicular to q. We can therefore use them as directions for the three axes that we need for the tangent space at q, and they vary continuously in a beautiful way.

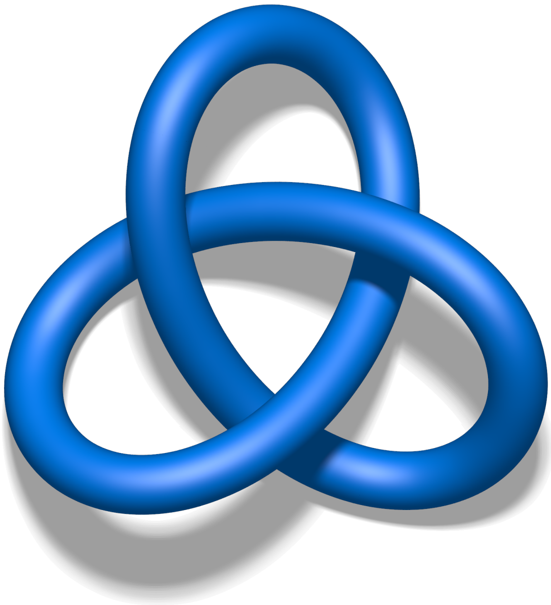
Can we define nice notions of "multiplication" for points on spheres in other dimensions? If we want to use a standard number system, then the ones we have available are the real numbers (which don't actually help for this problem), the complex numbers (which give us the easy result that we can choose a direction at each point on a circle – just go round the circle and choose the forwards direction at each point), the quaternions and the *octonions* or *Cayley numbers*. A theorem of Hurwitz, proved in 1898, states that these are the only such systems. (I won't say precisely what this means.) The octonions are eight-dimensional numbers, so the unit octonions form a seven-dimensional sphere, and they can indeed be used to show that a seven-dimensional sphere is parallelizable.

Now the fact that we don't have convenient number systems around in the other dimensions does not by itself show that spheres in those other dimensions are not parallelizable: there might be some other method for choosing directions for the arrows. However, Milnor, building on work of Raoul Bott, showed in 1958 that 1, 3 and 7 were in fact the only dimensions for which a sphere is parallelizable, a result that I have seen described as "magical". This result was also obtained independently by Hirzebruch and Kervaire.

Incidentally, the hairy ball theorem is of interest here for two other reasons. First, Milnor came up with a surprising and beautiful new proof of the theorem in 1978 – surprising because the proof used multivariable calculus and not the tools of algebraic topology that had hitherto appeared to be essential, and beautiful because Milnor's proofs are always beautiful. Secondly, it gives some hint of the kind of reason that a homeomorphism between a sphere and an exotic sphere cannot have its kinks ironed out. If you take a hairy ball and the hair is bunched up somewhere, then you can brush that part to smooth it out, but the result will be to push the bunch to somewhere else rather than to get rid of it completely.

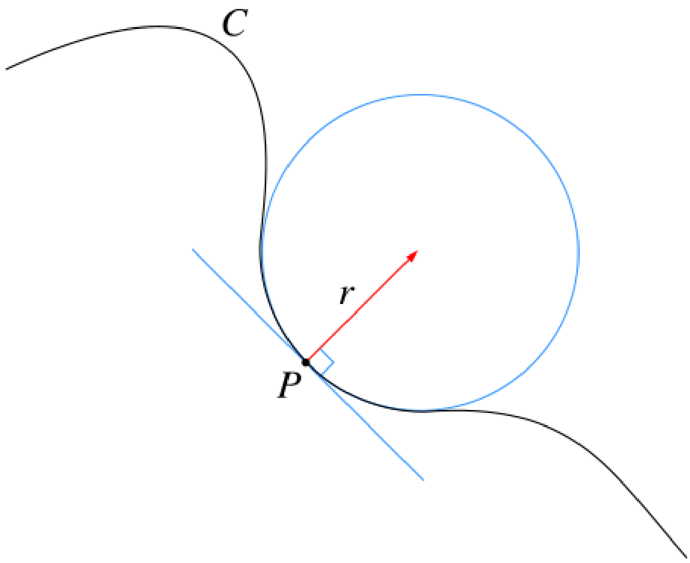
4.3. **How curved must a knot be?** Another beautiful geometrical result was proved by Milnor at the tender age of 19. It provides an answer to the following question. Suppose you take a curve in three dimensions that starts and ends in the same place and forms a knot. How curved must it be?

Of course, so far this is not a very precise question, but to get some intuition for it, consider the following picture, which shows the *trefoil knot*.



It is clear that in order for the curve to be knotted, it has to double back on itself rather more than a circle would. In fact, if you imagine looking at this knot from a point right in the middle, you can see that there is a sense in which the curve "goes round twice". And this, it turns out, is a highly relevant observation.

How can we measure the amount that a curve is curved? One way is to use the concept of the *radius of curvature*. The radius of curvature at a point P on a curve C is the radius of the circle that best approximates the curve near P, as the following diagram illustrates.



The *curvature* at P is the reciprocal of the radius of curvature: thus, the bigger the radius of curvature, the smaller the curvature and vice versa. This makes sense, since the more rapidly the curve is curving round, the smaller the approximating circle will be, and thus the bigger the curvature. The *total curvature* of a curve is what you get when you sum up (or, more precisely, integrate) the curvature.

If the curve is itself a circle of radius r, then the approximating circle at each point is that very circle, so the radius of curvature at each point is r and the curvature is $1/r$. Since the length of the curve is $2\pi r$, it follows that the total curvature is $2\pi r$ multiplied by $1/r$, which is 2π. (In fact, the same is true for any closed curve in the plane, if we interpret the curvature as negative when the approximating circle lies on the outside.) If we believe the intuition that a knot has in some sense to "go round twice" then we might expect the total curvature of a knot to have to be more than 4π. (I say "more than" rather than "at least" because a knotted curve cannot actually stay in a single plane.) This expectation was proved correct by Fáry and Milnor in independent work (Fáry in 1949 and Milnor in 1950).

I cannot resist including the following illustration, which shows Mjölnir, the hammer of Thor. Its relevance is threefold: "Mjölnir" is almost an anagram of "Milnor", Thor is a major god in Norse mythology, and if you look closely at this particular rendition of Mjölnir you will see two trefoil knots.

5. Growth in groups

Two years ago the Abel Prize was awarded to Mikhail Gromov, who, like Milnor, has had a deep influence on many different areas of mathematics. One of Gromov's most famous results is his *polynomial growth theorem*, which generalizes yet another beautiful theorem of Milnor, this one proved with Wolf in 1968.

I do not want to say here what a group is, but to give some idea of what the theorem says I shall give two examples, both of groups "generated by the two symbols a and b". In both cases, the group consists of sequences of terms, each of which is either a, b, a^{-1} and b^{-1}. These sequences are called "words". For example, $aba^{-1}bbbaab^{-1}a$ is a typical word.

In addition, there is a very basic rule for combining words: given two words, write the first one and then the second. For example, if I combine the words aba and aab^{-1} I get the word $abaaab^{-1}$. What makes the subject interesting is that we also have rules that tell us that we want to regard some words as being equal, and these rules vary from group to group.

The minimal rule we are allowed to apply is a rule that whenever you have the symbols a and a^{-1} next to each other, then the word is considered the same as the word that would result if you cancelled those symbols out. For example,

$$ab^{-1}aa^{-1}bbba = ab^{-1}bbba = abba,$$

because we can cancel the aa^{-1} and then the $b^{-1}b$. It is not hard to show that if this is the only rule, then every word in the group has a *standard form*: it can be written in precisely one way as a sequence of as, bs, a^{-1}s and b^{-1}s in such a way that no a is next to an a^{-1} and no b is next to a b^{-1}. (To do this, you just repeatedly cancel as in the example above.) This group is called the *free group on two generators*.

A much more drastic rule, in the sense that it makes many more words equal, is to say that the order of the symbols does not matter. Then we would count $aba^{-1}b$ as the same word as $aa^{-1}bb$, which in turn would be equal to bb (since the "inverse pairs" rule always applies in any group). In this case it is easy to see that every word can be written in precisely one way as a string of as or a^{-1}s followed by a string of bs or b^{-1}s. For instance, two typical words are $aaaaab^{-1}b^{-1}$ and $abbbbb$. Groups where the order of the symbols does not matter are called *Abelian*, after the same Niels-Henrik Abel whose name is attached to the prize that has just been announced.

The *growth rate* of a group is the function that tells you, for each n, how many distinct words there are of length at most n. For the Abelian group above, this is comparable to

n^2, since once you have decided whether you want a or a^{-1} and whether you want b or b^{-1}, all that is left is to choose two lengths that add up to at most n. Because n^2 is a polynomial in n, we say that this group has *polynomial growth.*

In the case of the free group, there are many more words of length at most n. In fact, there are exponentially many. To see this, note that if you want to build up a word in such a way that no cancellation is possible, there are four choices for the first letter (a, b, a^{-1} or b^{-1}) and three choices for each subsequent letter (all you have to do is avoid the one symbol that would cancel with the one before it). This means that the number of non-cancellable words of length n is 4.3^{n-1}, which grows exponentially fast as n grows.

It is quite easy to show that the growth rate in any Abelian group (not necessarily with just two generators) must be polynomial, and the same is true for a class of groups called *virtually nilpotent*, which are "almost Abelian" in a precise mathematical sense. Milnor and Wolf showed that for a wide class of groups, the *solvable* groups (so-called because they are closely related to the famous work of Abel and Galois on solving polynomial equations) the converse is also true: if a group has polynomial growth then it is virtually nilpotent. This is a notable fact because an apparently very flexible condition, polynomial growth, has a highly structural and algebraic consequence, that of being virtually nilpotent.

In 1981 Gromov proved the same result for *all* (finitely generated) groups, and not just the solvable ones.

6. Further contributions

I have by no means exhausted Milnor's work – indeed, there is no hope of my doing so. But let me finish by mentioning two other areas in which he has made major contributions.

6.1. **Algebraic K-theory.** A question that I did not address earlier is the important one of how Milnor was able to show that his exotic spheres really were exotic: how could he be sure that they were not diffeomorphic to ordinary spheres? In many situations, the best way to show that two objects are distinct is to find an *invariant* that distinguishes between them. What this means is that you find something that you can calculate in terms of the mathematical object in question, which should have the following two properties:

(i) if the objects are the same (in the sense of "sameness" that you are interested in), then the values of the invariant for those two objects are equal;

(ii) for the two objects you want to prove are distinct, the values of the invariant are different.

If you can do this, then you have indeed proved that the objects are distinct.

To give an example, it is not all that easy to prove rigorously that the trefoil knot is genuinely distinct from a circle, but people have discovered ways of associating polynomials with knots with the property that if you manipulate a knot without cutting it, then the polynomial does not change, and also with the property that the polynomials associated with the trefoil knot and the circle are distinct.

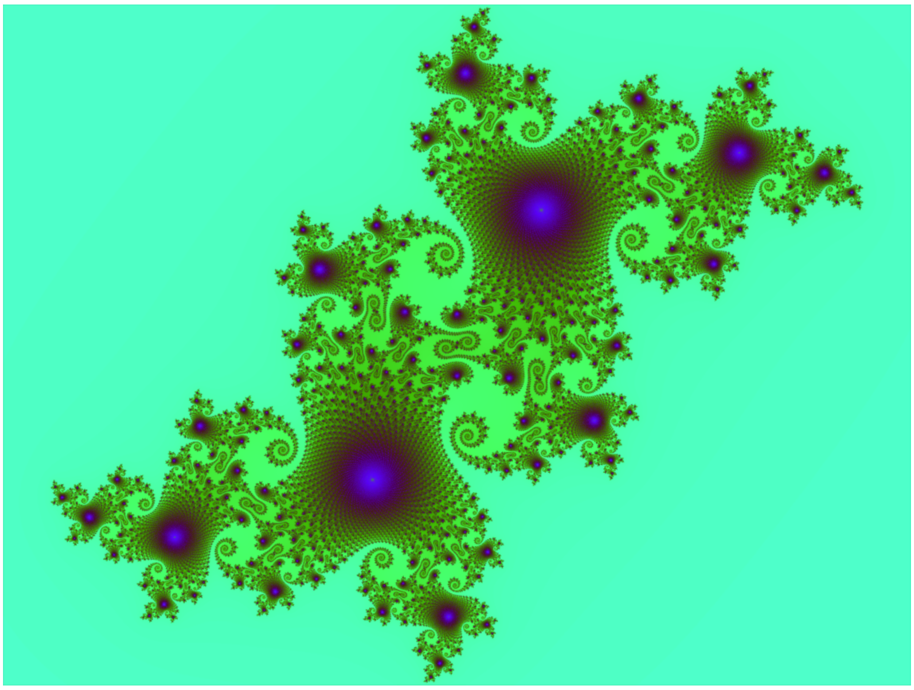
A very important class of invariants emerged in the 1950s and led to a part of mathematics known as K-theory. It was pioneered by Grothendieck in the context of algebraic geometry, and later by Atiyah and Hirzebruch in topology.

A fundamental concept in topology is that of a *homology class.* Without going into details, we can think of this as something like a lower-dimensional submanifold of a manifold, except that we regard two such submanifolds as "essentially the same" if one can be continuously moved to the other. (This is a slightly misleading way of putting it, but will do for the purposes of this discussion.) In algebraic geometry, one likes to deal just with objects that arise as sets of solutions of polynomial equations: let us call these *algebraic sets.* The question then arises of whether we can make sense of topological concepts such as homology classes in algebraic terms. In particular, can we choose our lower-dimensional submanifolds to be algebraic sets if the original manifold is an algebraic set? Several questions of this general flavour are still open and are amongst the deepest in mathematics.

In the 1960s, it became clear that there ought to be an algebraic version of K-theory that would be helpful for some of these questions. However, it was far from clear how to define some of the key concepts that would be required in this theory. What was being looked for was a sequence of groups $K_0, K_1, K_2, \dots$ associated with a ring A. Before Milnor's work, it was understood how to define the groups $K_0(A)$ and $K_1(A)$. Milnor found the right definition of $K_2(A)$, and proposed definitions of all the higher K-groups. Eventually, Quillen found the right definitions for all the K-groups.

As part of this work, Milnor formulated a conjecture that became sufficiently important that Vladimir Voevodsky was awarded a Fields medal in 2002 for solving it. (It is notable just how dense this area is in Fields medallists: Grothendieck, Atiyah and Quillen were all Fields medallists as well.)

6.2. **Holomorphic dynamics.** Since about the mid-1980s, Milnor has been working in the field of holomorphic dynamics, the area of mathematics that concerns iterating maps on the complex numbers, on more general Riemann surfaces (spaces that look locally like the complex numbers rather as a manifold looks locally like Euclidean space), and on higher-dimensional complex structures. Holomorphic dynamics is the branch of mathematics that leads to pictures like this.



The results and techniques are substantially different from the results and techniques of differential topology, but Milnor has an excellent way of dealing with this kind of problem: he learns what he can from the experts, reworks it, and writes a classic book on the subject, in this case Dynamics In One Complex Variable.

He has also been working on dynamics in more than one complex variable. Introducing more variables makes a big difference and makes many of the results much harder. Milnor has been a key figure in this area.

7. Summary

I intended this description of Milnor's work to be much shorter, but such is the sheer quantity, variety and beauty of what he has done that I found myself unable to stick to my plan. He is truly one of the greats, a mathematician who has hugely enhanced the subject and inspired many others. The award of the Abel Prize to John Milnor will be welcomed by mathematicians the world over.