Mathematics as a basic science

Michael Atiyah

IT is a great honour and pleasure for me to deliver the first Rajiv Gandhi Science Lecture, particularly in the presence of His Excellency the Governor of Karnataka and of Mrs Sonia Gandhi.

As President of the Royal Society, which numbers many distinguished Indian scientists among its Fellows, I am delighted to be here addressing an audience with an interest in science. Rajiv Gandhi, like his grandfather Jawaharlal Nehru, was a strong supporter of science and both of these Prime Ministers of India studied at my college in Cambridge. This gave me an added reason for accepting your kind invitation.

We all know the fundamental role that science plays in our lives. It underpins everything in modern society; engineering, medicine even agriculture are now built on scientific foundations and this process is accelerating. Two of the most dramatic examples are provided by the computer revolution of the past decades, based on the incredible miniaturization of components, and the dawn of molecular genetics based on the understanding and manipulation of DNA.

The whole scene is both daunting and bewildering; it is beyond the grasp of any one person to appreciate fully what is happening. But there are two key points which should be stressed. The first is the essential unity of science. No part of science, however specialized, thrives in isolation. Techniques and ideas from one field overflow into quite different areas. For example, our understanding of molecular genetics would have been impossible without the use of X-ray crystallography, essentially a technique based on physics and developed by chemists. Moreover interpreting the data that crystallographers observe requires some sophisticated mathematics. Finally the vast amount of data could not be handled without the incredible power and speed of *modern computers.*

Modern science is a massive and complex structure of interconnecting parts. Moreover, the links are frequently unexpected and unpredictable.

The second important point is the relation between basic science and applied science. Much effort has gone into trying to define these two terms and to introduce intermediate categories such as 'strategic science', but I will not enter here into these subtleties. We can broadly define the aims of scientists as that of understanding the natural world (in the broadest sense) and of using that understanding for the benefit of mankind. Knowledge and use are the dual objectives. Some parts of science are, at any given moment, of more immediate use than others but, taking the longer view, there is no such thing as useless science.

In any structure some components are nearer the foundation and some are nearer the roof. In a literal sense basic science refers to those parts of science which support the superstructure. In that sense one can reasonably argue that mathematics is the most basic of the sciences. Because of its antiquity and its logical nature mathematics is, in varying degrees, indispensable to all the sciences. It provides the framework and language in which much of science is formulated. It provides precision and clarity, enabling other scientists to establish laws and deduce consequences.

Let me give a few illustrations which demonstrate the wide range over which mathematics operates, and the unexpected ways in which it can become applied.

Perhaps the oldest and deepest applications of mathematics relate to physics, from Newton's time onwards. The laws of electricity and magnetism provide an impressive example. Based on the experimental work of Faraday, Clerk Maxwell in 1865 formulated the mathematical equations which now bear his name. These equations describe in a compact form the inter-relation between electricity and magnetism. They led in due course to the discovery of radio waves and they also encompass the theory of light. No mathematical equation has ever had greater practical importance. We might justifiably say that modern society has this equation as its foundation stone.

I have already mentioned the role of mathematics in crystallography. Here the relevant mathematics was developed by the French mathematician Fourier in the early 19th century, in connection with the study of heat. It has proved fundamental in all branches of physics involving the study of waves as in water, light or sound. A somewhat analogous but more recent story involves the development of the CAT scanner. This involves the use of X-rays in medical practice to locate the 3-dimensional position and density of objects in the human body. One set of X-rays gives only a 2-dimensional projection but, by using X-rays in different directions, more information is obtained and one would like to reconstruct a fully 3-dimensional picture from a number of 2-dimensional ones. Fortunately, the mathematics of this has been studied, purely for its own sake, by Radon several decades earlier and an elegant solution had been obtained which could now unexpectedly be put to use.

Text of the Rajiv Gandhi Science Lecture given on 28 October 1993 in Bangalore

Michael Atiyah is President, The Royal Society, 6 Carlton House Terrace, London SW1Y 5AG, UK.

I have referred to the computer revolution, one of the significant technological events of our time. In fact, the early pioneers in computing, such as Babbage in the 19th century, Turing and Von Neumann in the 20th century, were all mathematicians. There is an obvious analogy between the logical steps of a mathematical argument and the operation of digital computers. Mathematical logic was once regarded as the most abstruse branch of mathematics, more akin to philosophy, as in the work of Bertrand Russell. Now students with Ph D.'s in logic are snapped up by computer companies to help them design the software of the future.

Another example of totally unexpected applications of pure mathematics is provided by the recent use of prime numbers in the construction of codes. Anyone can see at a glance that 15 is 5 times 3 but factorizing a number of say 50 digits is a different matter, particularly if its prime factors are themselves large. Neither theory nor computers are of much help and so, based on this fact, one can construct simple unbreakable codes.

Users of such codes include banks and other financial institutions so that, if utility is measured in pounds, prime numbers are certainly being put to profitable use.

The famous British mathematician G H. Hardy gloried in the fact that the theory of Numbers, his own specialty, would never be put to practical use. He must be turning in his grave.

Number Theory for its own sake, as a great intellectual challenge has a long history, particularly here in India. Already in the 7th century Brahmagupta made important contributions to what is now known (incorrectly) as Pell's Equation.

This equation is of the simple form

$$x^2 - Ny^2 = 1,$$

where N is an integer. The difficulty is that we require the unknowns x and y also to be integers.

For example when $N = 2$ it is easy to find the solution $x = 3$, $y = 2$. But for $N = 61$ it turns out that the smallest solution is

$$x = 1766319049$$

$$y = 226153980$$

indicating the unexpected difficulty of the problem.

More recently, of course, and linked to G. H. Hardy, was the remarkable case of Ramanujan, a self-taught genius who produced profound results by mysterious methods. Ramanujan contributed to many aspects of the Theory of Numbers, but he was specially fond of infinite series with remarkable arithmetic properties. There is, for example, the function named after him which enumerates the coefficients of a certain infinite product:

$$x\ (1-x)^{24}\ (1-x^2)^{24}\ (1-x^3)^{24}\ldots$$

$$= x - 24x^2 + 252x^3 - 1472x^4 + 4830x^5 \ldots.$$

Srinivasa Ramanujan

Ramanujan was rightly fascinated by these large interesting integers and he made remarkable predictions about them which have subsequently been verified.

This year number theory happened to hit the headlines and found its way on to the front pages of newspapers all round the world. I am referring to the solution of 'Fermat's Last Theorem', which was announced by Andrew Wiles at a lecture in Cambridge last June. This finally solved the most famous problem in mathematics – a problem formulated by Fermat over 300 years ago and which has resisted the efforts of the best mathematicians over the intervening centuries.

Popular interest in this story was enhanced by two additional facts. First, the problem is very simple to state, so that it can be understood by the man in the street. Second, Fermat wrote in the margin of his book a cryptic note saying that he had discovered a marvellous proof but the margin was too small to write it down.

Andrew Wiles

Pierre de Fermat

Viscount Brouncker

This remark has bemused and perhaps irritated his successors but it has added to the fame of the problem.

There are also other factors which make Fermat's Last Theorem of special interest to the President of the Royal Society.

Fermat was by profession a lawyer, working in the south of France, but mathematics was his consuming interest. He corresponded with some of the leading scholars in Europe. As was customary at the time, he would pose problems and issue an open challenge to anyone to produce solutions.

In fact, one of the problems asked for a general method to solve Pell's equation, a problem which unbeknown to his contemporaries had, as I have already mentioned, been studied and essentially solved by Indian Mathematicians many centuries earlier.

Among those who sparred with Fermat, responding to his challenges, were two of the leading English mathematicians of the time, John Wallis and Viscount Brouncker. Brouncker subsequently became the first President of the Royal Society; so it seems appropriate that 330 years later, when a mathematician is again President of the Society, you should hear the end of the story.

You will be relieved to hear that I will spare you the technical details. Fortunately, they require more expertise than I possess. However, I would like to deal with the problem on a grand scale, indicating how it has been linked with the major events of the past 300 years. It is a story which contains many lessons relevant to our time and to fields other than mathematics, emphasizing the unexpected way in which science develops and the intricate interconnection of its component parts. In a spectacular way it exemplifies the long time-scales that may be needed.

The problem starts with the well-known observation that there are right-angled triangles whose sides (in some unit of measurement) have integer lengths. The best known example is the (3, 4, 5) triangle and the next is (5, 12, 13). By the famous theorem of Pythagoras the sides (a, b, c) of such a triangle have to satisfy the equation

$$a^2 + b^2 = c^2.$$

It was already known to the Greeks that there were infinitely many integer solutions to this equation and so infinitely many right-angled triangles with integer sides. If p, q are any integers, a little elementary algebra shows that

$$a = p^2 - q^2, \quad b = 2pq, \quad c = p^2 + q^2$$

provide solutions of the Pythagorean equation. Fermat asked: What happens if we replace squares by cubes? In other words, can we find positive integer solutions of the equation

$$a^3 + b^3 = c^3 ?$$

More generally, he asked the same question for an arbitrary integer exponent n. Does the equation

$$a^n + b^n = c^n$$

have any positive integer solutions (a, b, c)? His 'Last Theorem' asserts that, for n greater than 2, the answer is no.

In the 17th, 18th and 19th centuries this problem attracted the attention of the most famous mathematicians, spurred on no doubt by prizes offered by the French Academy. For $n = 4$, a proof essentially that of Fermat himself, was published in 1676. A century later Euler disposed of the more difficult case $n = 3$. By 1840 Legendre and Dirichlet had dealt with $n = 5$ and $n = 7$. The proofs of these special cases depended in part on the solution of Pell's equation and this related it to Fermat's other interests.

A little later, Kummer claimed to have a general proof for all exponents. Unfortunately, although Kummer's proof was a great step forward, there was a subtle fallacy in the argument. Analysing this error (and proving Fermat's Theorem for many values of n all at once) opened up a vast new field, and led Kummer to develop deep new ideas which have had widespread impact on many branches of mathematics. Kummer's error is quite easy to explain. We all know that ordinary integers can be factorized into prime factors (e.g. $24 = 2^3 \times 3$). Moreover, this factorization is unique, an apparently rather obvious fact which is not usually stressed. There are 'generalized integers' which share similar properties: e.g. expressions $a + b\sqrt{2}$ with a, b ordinary integers or $a + b\sqrt{-1}$. These last occur naturally in connection with the Pythagorean equation ($n = 2$) because of the identity

$$a^2 + b^2 = (a+b\sqrt{-1})\,(a-b\sqrt{-1}).$$

Perhaps at this stage I should comment on $\sqrt{-1}$ and 'imaginary numbers'. When these were first introduced, purely formally, as in the above equation, there was much controversy over their real meaning.' But, as mathematics evolved, the utility of imaginary numbers was increasingly recognized. They are a great help in connection with Fourier's theory, mentioned earlier, and are now extensively used by electrical engineers. If immediate practical use had been a criterion at the time, reaearch on imaginary numbers would certainly not have been funded!

Returning to our 'generalized integers', Kummer assumed implicitly that their factorization into prime factors was always unique. This turns out to be a subtle property, sometimes true and sometimes false, depending on the class of integers. For example, it is false for expressions $a + b\sqrt{-5}$ with a, b ordinary integers.

Kummer's brave and fruitful, if inconclusive, attempt was rewarded by the French Academy with a prize of 3000 francs. In 1908 the Academy in Göttingen received a private benefaction offering a prize of 100,000 Deutschemark for a general solution of the Fermat problem. Recognizing its difficulty, the offer had a generous time limit: solutions had to be received by 13 September 2007.

Despite these incentives, there has been, until this year, no fundamental progress, although computer calculations had, through a combination of theory and brute force, disposed of more and more values of the exponent n. Needless to say, the apparent simplicity of Fermat's Theorem has attracted hordes of amateurs who delude themselves into thinking that they have found a proof. There must now be enough false proofs to fill a library.

The Fermat equation is, of course, a very special equation, and however beautiful and fascinating it may be, mathematicians like to understand things in greater generality. We can obviously consider more complicated equations and ask for integer solutions. This branch of the Theory of Numbers is called Diophantine (after the Greek mathematician Diophantus), and Fermat and Kummer are two of the principal figures in its recent history.

Of course, an equation can be viewed as defining a curve. For example, replacing a/c and b/c by x and y in the Pythagorean equation, we recognize the usual equation of a circle

$$x^2 + y^2 = 1.$$

Integer triples (a, b, c) now correspond to points on the circle with rational fraction coordinates (x, y): we call them rational points. Of course, as known to the Greeks, not all numbers are rational (fractions). For example, $\sqrt{2} = 1.414\ldots$ can only be represented by an infinite (non-recurrent) decimal.

The geometric point of view is a powerful one and, over the past century, there has been much progress in looking for 'rational points' on algebraic curves. It turns out that there is a crude, but fundamental, labelling of curves by an integer called the genus g (for Fermat curves this is related to the exponent n). If $g = 0$ the answer is easy: if there is one rational point there are infinitely many and there is a simple rule to describe them all (as with the Fermat curve for $n = 2$). If $g \geq 2$ rational points are rare and a famous conjecture of Mordell proved a few years ago by Faltings asserts that the number of rational points is always finite. This includes all Fermat curves with $n \geq 4$ and is a step on the road to showing that the number is actually zero.

The case $g = 1$ (corresponding to the Fermat curve with $n = 3$) is the most interesting and such curves are called elliptic, being related to elliptic functions. Here the numbers of rational points may be finite or infinite. They have an intricate structure but they are very difficult to find and enumerate. As a result, a great deal of effort by algebraic geometers and number theorists has been devoted to elliptic curves. Some remarkable conjectures of Birch and Swinnerton-Dyer in 1960, concerning rational points on elliptic curves, have been central to research since that time and may, in due course, acquire a status comparable to Fermat's Theorem.

Now curves can be described in two complementary ways, either by an equation or by a parametrization. For example the curve whose equation is

$y^2 = x^3$

can be parametrized by using a parameter t with

$x = t^2, \quad y = t^3.$

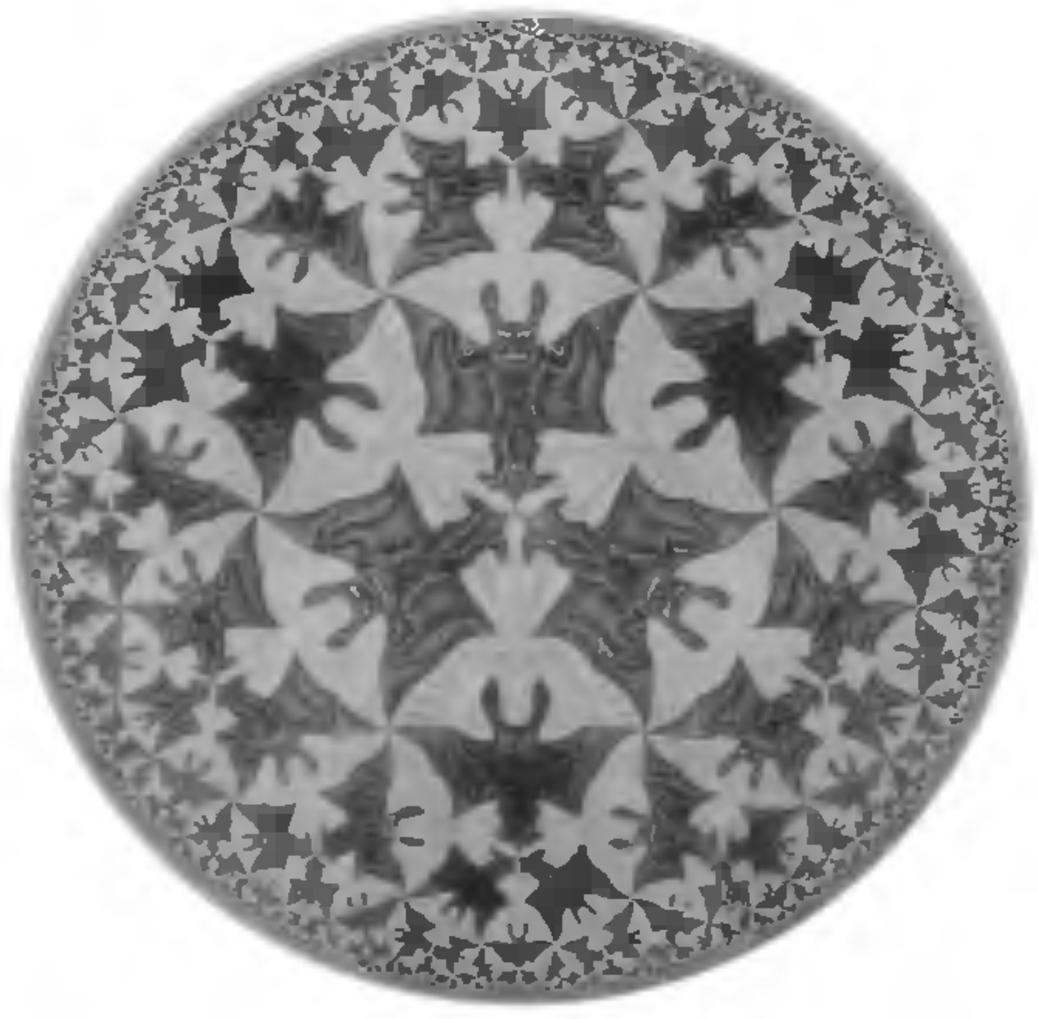
Sometimes a point corresponds to a unique value of the parameter (as with this example) but in other cases many values of t may yield the same point, i.e. there is some redundancy in the parametrization. For example, if we used a thread round a circle (of unit length) then t and $t + n$ (with n an integer) give the same point. A more complicated example arises when the parameter t and

$t' = (\alpha t + \beta)/(\gamma t + \delta)$

with $\alpha, \beta, \gamma, \delta$ integers, might define the same point. The transformation from t to t' is called *modular* and a curve with a modular parametrization is called a modular curve. It turns out that modular curves are much easier to understand and, as you might expect, a lot is known about rational points on them. In particular this holds for elliptic modular curves.

Let me digress for a moment to talk a bit about modular transformations. These turn up in many parts of mathematics and they have some fascinating geometry attached to them. This is best illustrated by one of Escher's pictures, which also illustrates with artistic licence the aesthetic component in mathematics. The modular transformations are symmetries of the picture, they move each 'cell' on to another one.

The fact that cells near the boundary look smaller than those near the centre is a distortion analogous to the familiar distortion in geographical maps, in which the arctic and antarctic regions are unduly magnified. Notice however that the distortion in the Escher picture has the opposite effect of shrinking, rather than expanding, as we move to the outer region. In fact Escher's picture is a model of the hyperbolic or non-Euclidean plane and incidentally is also closely related to imaginary numbers. The discovery that there were other geometries besides that of Euclid (in which the angles of a triangle need not add up to 180°) was one of the major events in mathematical history. In turn it led to more general curved geometries of the type which were ultimately used by Einstein in his theory of General Relativity. So modular transformations are firmly embedded in the history of mathematics and physics. Moreover modular transformations and their number-theoretical implications lie behind much of Ramanujan's work and in particular are the key to the Ramanujan function.



Angel–Devil (Escher)

Number theorists therefore had two classes of elliptic curves before them. Those given by equations which were difficult to study and those which were modular and better understood. Wouldn't it be wonderful, they said, if all elliptic curves (given by any equation) were in fact modular? This apparently naive idea was first put forward, about 40 years ago, by the Japanese mathematician Taniyama. It was subsequently refined and shown not to be so naive by many other mathematicians. In fact it was shown to fit into a very general theory put forward by Robert Langlands. This Langlands' programme has all the same ingredients: symmetry, equations, rational points, but now in a much more comprehensive framework.

Let me just say a few words about symmetry. The study of symmetry in mathematics began in fact with the pioneering work of the young French mathematician Galois (who died in a duel at the age of 21). Galois was studying the symmetry in solutions of equations: for example the symmetry between $+\sqrt{2}$ and $-\sqrt{2}$ as solutions of $x^2 = 2$. I have already mentioned geometric symmetries as in non-Euclidean geometry. In general, symmetry provides one of the unifying and simplifying features throughout mathematics. It also permeates physics, from the symmetries of crystals (a subject familiar to the great Indian scientist C. V. Raman, whose museum I visited today) to the symmetries of relativity theory and fundamental particles. We now understand that the basic conservation laws of energy, momentum or electric charge are all manifestations of underlying symmetries. Digging deeper into the nature of matter now involves a constant search for 'hidden' symmetries.

Moreover, the Langlands' programme borrows some of its key concepts from that of quantum theory. You might think it highly implausible that physics should have anything to contribute to the theory of numbers. But in fact quantum physics does deal with discrete quantities, like energy levels, and there are certainly situations in which enumerating integer solutions is related to quantum-mechanical problems.

But I am straying from my point. I was just explaining that, as part of the general Langlands programme, the naive conjecture of Taniyama takes its natural place.

Andrew Wiles announced this summer the general truth of Taniyama's conjecture that (almost) all elliptic curves are modular. This is a very important theorem and typical of the thrust of contemporary mathematics – to establish very general results as part of a big picture. Of course one hopes such general results will have specific pay-offs in various special cases but what, you might ask, has this to do with Fermat's Last Theorem which (for $n \geq 3$) was not about elliptic curves but apparently about curves of higher genus?

The answer lies in a very clever observation by Frey, put on a rigorous footing shortly afterwards by Ribet. The observation went as follows: suppose we have a solution, in positive integers, (a, b, c) of the Fermat equation of exponent n. Consider the *elliptic* curve

$$y^2 = x(x - a^n)(x + b^n).$$

Surprisingly it turns out that one can show this curve is *not* modular.* This contradicts Taniyama's conjecture (now established by Wiles) and hence our hypothetical solution of the Fermat equation cannot exist. In other words Fermat's Last Theorem has been proved

This last twist in the story carries a lesson for all of us in mathematics or science. A direct onslaught on a difficult problem may get stuck but, if we persevere with a general investigation of related matters, trying to understand the fundamentals then, in due course, we may be able to solve our earlier problem as an incidental by-product.

So a problem formulated before the foundation of the Royal Society has finally been solved. At least we hope so! Until a mathematical proof of this degree of sophistication has been completely written down and subjected to critical examination it may still turn out to contain crucial gaps. The example of Kummer provides a warning but also some consolation. Even if Wiles' proof turns out to be incomplete it will certainly have shed new light on the problem and opened up further doors. Perhaps a problem that constantly eludes us, but stimulates new ideas and techniques is preferable to one that can be finally disposed of.

Throughout these three centuries Fermat's Last Theorem has challenged mathematicians and has acted as a test of mathematical technique and virtuosity. Wiles' solution is an outcome of many theories that have been built up since Fermat's time. Totally new ingredients such as symmetry, which I have mentioned, and topology, which I have not, have contributed to the whole framework. Mathematicians now have extremely sophisticated and refined tools at their disposal and all of these are involved in the proof of the Fermat Theorem. You are all familiar with the sophisticated apparatus that experimental scientists now use: physicists, chemists and biologists. These enable the modern scientist to tackle questions that would be impossible to answer with more primitive equipment. The situation in mathematics is similar: we have a lot of elaborate machinery but in our case it consists of pure theory: ideas, techniques and formulae. You could call it software, but that hardly does it justice.

From what I have said, you will not be surprised to know that most mathematicians believe that Fermat was deluding himself when he made his cryptic claim in the margin of his book. He probably fell into a similar trap to Kummer, or perhaps he thought about the first cases of $n = 3, 4$ and assumed the rest would follow. It is, of course, still possible that some young genius will come along with a clever, new and direct proof of Fermat's Theorem – conceivably even one that Fermat could understand. This would be admirable and surprising, but it would in no way replace the great edifice that modern mathematics has erected and which Wiles has exploited. The aim of mathematics is to develop general theories, to provide understanding and unity – solving individual problems, however elegant and historic, is incidental and is merely a test of the genuine power and applicability of the theory that has been developed.

In recounting this fascinating story I have tried to draw out some of the lessons which are implicit in it on the nature of scientific progress. The time-scale of 300 years is unprecedented in modern times but it does emphasize the virtues of patience. Many of the problems we face will require decades if not centuries and we cannot guarantee solutions tomorrow. I have also indicated that Fermat's Theorem eventually emerged from a general framework of ideas that had roots in many different parts of mathematics and physics. I am a strong believer in the essential unity of science and Fermat's Theorem is a good illustration.

In this day of financial incentives, it is also salutory to note that, despite handsome prizes being repeatedly offered for the proof of Fermat's Theorem, money played no significant part in the end. The prize of 100,000 pre-World-War-One Deutschemark is hardly a major inducement.

I like to reflect on what would have happened had King Charles II, who founded the Royal Society, decided, perhaps on the advice of Viscount Brouncker, that national prestige (or Anglo-French politics) required a British solution of the Fermat problem. What kind of targeted research programme would the Royal Society have been told to initiate? Perhaps Newton would have been instructed not to waste his time on the calculus but to concentrate on algebra? A decade or so later, with no progress to report, the Royal Society would have been restructured and merged with the Mint, so as to give Number Theorists a more practical bent. I leave the rest to your imagination!

*At least for n an odd prime, but this is sufficient for the Fermat Theorem